

Fri Mar 14 14:49:32 2003

us-09-674-237a-3.ra1

Page 1

GenCore version 5.1.4.p5.4578  
Copyright (C) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2003, 11:37:52 ; Search time 23 Seconds  
(without alignments)  
1551.739 Million cell updates/sec

Title: US-09-674-237A-3  
Perfect score: 6269  
Sequence: 1 MAQFPTFGSLDVAIVTE.....VGLFSPNVKLTMDPSQ 1213

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents, AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PTCTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2058.5	32.8	462	4 US-08-630-915A-38	Sequence 38, Appl
2	1388.5	22.1	509	4 US-08-630-915A-194	Sequence 194, App
3	816.5	13.0	248	4 US-08-630-915A-40	Sequence 40, Appl
4	486.5	7.8	896	1 US-08-095-737-2	Sequence 2, Appl
5	486.5	7.8	896	1 US-08-480-145-2	Sequence 2, Appl
6	486.5	7.8	896	2 US-08-477-389-2	Sequence 4, Appl
7	486	7.8	897	1 US-08-095-737-4	Sequence 4, Appl
8	486	7.8	897	1 US-08-480-145-4	Sequence 4, Appl
9	486	7.8	897	2 US-08-477-389-4	Sequence 4, Appl
10	382	6.1	1162	4 US-08-728-333A-2	Sequence 2, Appl
11	382	6.1	1162	4 US-09-298-568-2	Sequence 2, Appl
12	381	6.1	1898	2 US-08-056-200-94	Sequence 94, Appl
13	381	6.1	1898	2 US-08-800-644-84	Sequence 2, Appl
14	333	5.3	553	1 US-08-475-894-2	Sequence 2, Appl
15	333	5.3	553	1 US-08-484-709-2	Sequence 2, Appl
16	333	5.3	553	4 US-08-474-697-2	Sequence 2, Appl
17	333	5.3	553	4 US-08-474-697-2	Sequence 2, Appl
18	325	5.2	659	4 US-08-630-915A-133	Sequence 133, App
19	322	5.1	62	4 US-08-630-915A-135	Sequence 135, App
20	314	5.0	62	4 US-08-630-915A-134	Sequence 134, App
21	309	4.9	56	4 US-08-466-390-4	Sequence 4, Appl
22	297	4.7	2101	1 US-08-470-950-4	Sequence 4, Appl
23	297	4.7	2101	1 US-08-467-781-4	Sequence 4, Appl
24	297	4.7	2101	1 US-08-483-924-4	Sequence 4, Appl
25	297	4.7	2101	2 US-09-452-294-1	Sequence 1, Appl
26	297	4.7	2101	2 US-09-452-294-1	Sequence 1, Appl
27	296	4.7	57	4 US-08-630-915A-136	Sequence 136, App

28	296	4.7	2101	1 US-08-195-487-4	Sequence 4, Appl
29	296	4.7	2101	5 PCR-US93-06160-4	Sequence 2, Appl
30	283	4.5	1805	1 US-07-853-913-2	Sequence 2, Appl
31	281	4.5	683	6 Patent No. 5210183	Sequence 30, Appl
32	281	4.5	788	4 US-08-630-915A-30	Sequence 2, Appl
33	280.5	4.5	1375	4 US-09-722-139-2	Sequence 2, Appl
34	280.5	4.5	1375	4 US-09-721-832-2	Sequence 2, Appl
35	280.5	4.5	1375	4 US-09-393-569-2	Sequence 2, Appl
36	275	4.4	1360	4 US-08-685-871-2	Sequence 2, Appl
37	274.5	4.4	1354	3 US-08-533-306A-2	Sequence 2, Appl
38	274	4.4	576	2 US-08-742-923A-2	Sequence 2, Appl
39	274	4.4	576	2 US-08-533-306A-6	Sequence 6, Appl
40	265.5	4.2	816	2 US-08-742-923A-6	Sequence 6, Appl
41	265.5	4.2	816	2 US-08-533-306A-4	Sequence 4, Appl
42	265.5	4.2	885	2 US-08-742-923A-4	Sequence 4, Appl
43	265.5	4.2	62	4 US-08-630-915A-138	Sequence 138, App
44	263	4.1	955	1 US-08-006-676B-1	Sequence 1, Appl
45	259.5	4.1			

#### ALIGNMENTS

RESULT 1  
US-08-630-915A-38  
Sequence 38, Application US/08630915A  
Patent No. 6309820  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: HOFFMAN, No. 6309820h  
APPLICANT: KAY, Brian K.  
APPLICANT: FOWLER, Dana M.  
APPLICANT: MCCONNELL, Stephen J.  
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL  
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
TITLE OF INVENTION: USING SAME  
NUMBER OF SEQUENCES: 227  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/630,915A  
FILING DATE: 03-APR-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Mastrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-174  
TELEPHONE: (212) 790-9090  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 869-8864/9741  
TELEFAX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 38:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 462 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-630-915A-38  
Query Match 32.8%; Score 2058.5; DB 4; Length 462;  
Best Local Similarity 79.3%; Pred. No. 2e-138;  
Matches 391; Conservative 13; Mismatches 18; Indels 71; Gaps 1;

QY 721 FMSTTEKPLTISAQESKVVYVYALYFESRSHDEITTOPGDIYVWDESGTGEWLG 780  
 Db 41 PGPLOKGPPLTISAQENVKVVYVYALYFESRSHDEITTOPGDIYVWDESGTGEWLG 100  
 QY 781 ELKGTGWPANYAEKIPENEVPAKPYVDTLSAPAPKLALETAPLPTVTSSTTP 840  
 Db 101 ELKGTGWPANYAEKIPENEVPAKPYVDTLSAPAPKLALETAPLPTVTSSTTP 160  
 QY 841 NNMADPSTSSSENEKPEETDNDWTAAPSLTVPASGLRORSAFTPATATSSSPVL 900  
 Db 161 NNMADPSTSSSENEKPEETDNDWTAAPSLTVPASGLRORSAFTPATATSSSPVL 220  
 QY 901 GQGEKVEGLQAQALYPMRAKDNHNFNSKSDYITVLEQDDMMWGEVQOGKMPKSYK 960  
 Db 221 GQGEKVEGLQAQALYPMRAKDNHNFNSKSDYITVLEQDDMMWGEVQOGKMPKSYK 280  
 QY 961 LIGPVRKSTIDTGPTEPSALKRVASPAKPAIPGEETIAMYTESSSEGDLTPQGD 1020  
 Db 281 LIGPVRKSTIDTGPTEPSALKRVASPAKPAIPGEETIAMYTESSSEGDLTPQGD 318  
 QY 1021 VIVVTKDGDWMTGTVDKSGVFPNSYVRLDSESGTAGTSLGKKPIAVIASYAA 1080  
 Db 319 -----EIAQVIASYTA 329  
 QY 1081 TGPEOLTLAPGOLLIIRKKNPGWMEGLQARQKROIGFNPANYVLLSPGSKITPTE 1140  
 Db 330 TGPEOLTLAPGOLLIIRKKNPGWMEGLQARQKROIGFNPANYVLLSPGSKITPTE 389  
 QY 1141 LPTAVOPAVCOVIGMYDTAQNDELAFSGOIINVLNKEDPMWKGVSQVGLFPSN 1200  
 Db 390 LPTAVOPAVCOVIGMYDTAQNDELAFSGOIINVLNKEDPMWKGVSQVGLFPSN 449  
 QY 1201 YVKLTMDPSQO 1213  
 Db 450 YVKLTMDPSQO 462

## RESULT 2

US-08-630-915A-194

Sequence 194, Application US/08630915A

Patent No. 6309820

GENERAL INFORMATION:

APPLICANT: SPARKS, Andrew B.

APPLICANT: HOFFMAN, No. 6309820h

APPLICANT: KAY, Brian K.

APPLICANT: FOWLES, Dana M.

TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL

TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND

NUMBER OF SEQUENCES: 227

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie &amp; Edmonds LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/630,915A

FILING DATE: 03-APR-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Mirock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 1101-174

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090  
 TELEFAX: (212) 869-8864/9741  
 TELEX: 66141 PENNIE  
 INFORMATION FOR SEQ ID NO: 194:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 509 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: unknown  
 MOLECULE TYPE: peptide  
 US-08-630-915A-194

## Query Match

Best Local Similarity 22.1%; Score 138.5; DB 4; Length 509;

Matches 287; Conservative 85; Mismatches 124; Indels 91; Gaps 14;

QY 635 ERKSELEKQK-----EPAQRVQERDKQWLEHVQOEOPRPRKHEDRLKREDYRKE 690  
 Db 6 ERKSELEKQKLEBPAKAKAQKQENLWKENLKEE-----EKQRIQEKTEKQK 59  
 QY 691 AERARPEVQDQSRLEFHPHQPAPKATQAPWSTTEKGLTISAQSVVYVYALYPRE 750  
 Db 60 EERKAEEKQ-----RETASV-----LVNRYALYPRE 86  
 QY 751 SRSHDEITTOPGDIYVWDESGTGEWLGELKGTGWPANYAEKIP--ENEVTPAKP 808  
 Db 87 ARNHDEMSNSGDIQVDEKTVGEPGLYSGFQGNFGWPCNVYERKPSSENE----- 139  
 QY 809 VVDLSAPAPKLALETAPLPTVTSSTTPNNMADFSTWSSSENEKPEETDNDWTAAP 868  
 Db 140 -----KAVSPKAL-----LPTVLSLATS-----TSSEPLSSNDPASVTDYON--VS 180  
 QY 869 QPSLTVPASGLRORSAFTPATATSSSPVLGQGEKVEGLQAQALYPMRAKDNHNFN 928  
 Db 181 FSNLTIVTSMQ--KKAFTPTVSPG--SVSPHGOQGVENLKAQALCSWAKKDNHNFN 237  
 QY 929 KSDYITVLEQDDMMWGEVQOGKMPKSYVLLSGPVKSTSIDTGPTEPSALKRVAS 988  
 Db 238 KHDITVLEQDDMMWGEVQOGKMPKSYVLLSGPVKSTSIDTGPTEPSALKRVAS 289  
 QY 989 --PAKPAIPGEETIAMYTESSSEGDLTPQGDVIVVTKDGDWMTGTVDKSGVFPNS 1046  
 Db 290 KKFTSAVSVGEYIALYVSSVEPDLFTFGEIITVTDKDGWMTGSLGDRGIFPSN 349  
 QY 1047 YVRLKDSGSGTGTGKTSLGKKPEIAQVYASAAATGPEOLTLAPGOLLIIRKKNPGWME 1106  
 Db 350 YVRLKDSGSGTGTGKTSLGKKPEIAQVYASAAATGPEOLTLAPGOLLIIRKKNPGWME 409  
 QY 1107 GELQARQKROIGFNPANYVLLSPGSKITPTELPTAVOPAVCOVIGMYDTAQNDE 1166  
 Db 410 GELQARQKROIGFNPANYVLLSPGSKITPTELPTAVOPAVCOVIGMYDTAQNDE 462  
 QY 1167 LAFSGOIINVLNKEDPMWKGVSQVGLFPSNYYVLLTMDPSQO 1213  
 Db 463 LAFSGOIINVLNKEDPMWKGVSQVGLFPSNYYVLLTMDPSQO 509

## RESULT 3

US-08-630-915A-40

Sequence 40, Application US/08630915A

Patent No. 6309820

GENERAL INFORMATION:

APPLICANT: SPARKS, Andrew B.

APPLICANT: HOFFMAN, No. 6309820h

APPLICANT: KAY, Brian K.

APPLICANT: FOWLES, Dana M.

TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL

TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND

NUMBER OF SEQUENCES: 227

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie &amp; Edmonds LLP



STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10036-2711  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/630,915A  
 FILING DATE: 03-Apr-1996  
 CLASSIFICATION: 536  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Mistrock, S. Leslie  
 REGISTRATION NUMBER: 18,872  
 REFERENCE/DOCKET NUMBER: 1101-174  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 790-9090  
 TELEFAX: (212) 869-8864/9741  
 TELEX: 66141 PENNIE  
 INFORMATION FOR SEQ ID NO: 40:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 248 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: unknown  
 MOLECULE TYPE: peptide  
 US-08-630-915A-40

Query Match 13.0%; Score 816.5; DB 4; Length 248;  
 Best Local Similarity 58.4%; Pred. No. 1.6e-50; Indels 17; Gaps 3;  
 Matches 153; Conservative 38; Mismatches 54;

954 FPKSVKXLSGVRKSTSDTPTSPASLKRVAS--PAKPAIPGEFFIAMTYESESQ 1011  
 Db 2 FFAASTKIIIGSEVYK-----REPEALVAANVKRPTSAAYSVGEYIALYPSSEVP 53  
 Qy 1012 GDLTFOGDVIVTKKDGDMTGTIVGDKSGVFPSSNYVRLKSESGTIGKSGLSCKREI 1071  
 Db 54 GDLTTEGEBILVTKDDEWMTGSGIDRSGLFPSSNYVVRKPOESFGASKSGASNKREI 113  
 Qy 1072 AQVTSYATGEQELTLRAGLILRKKNPGMWEGELQARGKKRQIQMFANVYKLLSP 1131  
 Db 114 AQVTSYAVASGEQSLAPGQLILKKNTSQWQELQARGKKRQKQMFASVYKLLGP 173  
 Qy 1132 GTSKTPTELPPTAVQAVQVIGMVDYTAQNDDELAFSGQIIVLNKEDPDMMKGEVS 1191  
 Db 174 SSERATPAHP-----VCQVIAMVDVAANNDELSFGKGLIVNKKDDPDMMQGEIN 226  
 Qy 1192 GQVGLFPPSNVYKLTDDMPDQQ 1213  
 Db 227 GVTGLFPPSNVYKMTDSDPSQQ 248

RESULT 4  
 US-08-095-737-2  
 Sequence 2, Application US/08095737  
 Patent No. 5487979  
 GENERAL INFORMATION:  
 APPLICANT: DiFiore, Pier P  
 TITLE OF INVENTION: A Substrate for the Epidermal Growth  
 FACTOR RECEPTOR KINASE  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Knobbe, Martens, Olson & Bear  
 STREET: 620 Newport Center Drive, Sixteenth Floor  
 CITY: Newport Beach  
 STATE: California  
 COUNTRY: United States of America  
 ZIP: 92660

COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/095,737  
 FILING DATE: 19930722  
 CLASSIFICATION: 530  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Israelson, Ned A  
 REGISTRATION NUMBER: 29,655  
 REFERENCE/DOCKET NUMBER: NIH060.001A  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (619) 235-8550  
 TELEFAX: (619) 235-0176  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 896 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-095-737-2

Query Match 7.8%; Score 486.5; DB 1; Length 896;  
 Best Local Similarity 20.5%; Pred. No. 2.9e-26;  
 Matches 198; Conservative 150; Mismatches 273; Indels 345; Gaps 29;

Qy 15 MATTEERAKHQDFLSKPIGFTIGDARNFPPQSGLPQVLAQIYVALADNNDRMD 74  
 Db 122 WAKPBDKAKYDIAFDLSLPVFGSDKRVVLNSKLPVILGRVWELSDIDHMD 181  
 Qy 75 QVESIAMKLIKLGQYOLPSTLPVWKQPVASAPAFIGIGIASNPPLTAAVAPVM 134  
 Db 182 RDEFVAMFLVYALE-----XEPVYM 203  
 Qy 135 GSIPVVGMEPLVSVVPPAAPPPLANGAPPVLOPLPAFAPATMPKSSFSRSGSQL 194  
 Db 204 -----SLPPALVPP-----SKR 215  
 Qy 195 NTYKQAQSFVVASAPPAEMAVPOSSRLKYROLFNSHDKTMSGHLTGQARTIMQSSL 254  
 Db 216 KT-----WVSPAKAKYDEIFLTKTKMDGFPVSGLEVAEIFLKTGL 257  
 Qy 255 POAQLASIMNLSIDIDDGKLTAEFLAMHLIDVAM--SQQLPVPVLPPEYIPSEFRVRS 313  
 Db 258 PSTLLAHIMSLCOTKCKGSLKQDFALAFHLISQKLKIDPPIVLTPEMIPPS--DPA 314  
 Qy 314 GSGNSVISSSVDORLPEPSSDEQQPEKLPVTFEDKKREFFERGVSLELRQALLE 373  
 Db 315 SLQNTITGSSPV-----ADPSAIK-----ELDTLNNEIVD 344  
 Qy 374 QQRKEQRLAQLERABQERERQEQEAKRQLELEKOLEKORELROBEERKEIERR 433  
 Db 345 LQREK-----NNVEDDLKEKE-----DTIKQTSVQ--DLQDEVQRE 380  
 Qy 434 EAAKRELERQLEMERNRROELLNKNKEQEGTVLAKARKLTLELEALNDKKHQLBS 493  
 Db 381 NTNQKLOAQKQ-----QVQELLDE-----LDQKQALEB 410  
 Qy 494 KLQDIRLATQROEISTKSELRJAETHTLOQLQESQQLGRLIPKOLISQQLQ 553  
 Db 411 QLKVEKRKCAEAQLISLQ-----AELTQESQISTYEEELAK-----ARELSR 456  
 Qy 554 VQNSLHRDLSLTLRALKEKELARQOLRQLEDEVERETS-----KLOEIDVFNQQLK 607  
 Db 457 LQOETAR-----LEESVSGKQALEPLOQHLDQSQEISMQMKLMEMKDLNHNHSQLN 510  
 Qy 608 ELREIHS-----KQQLQKPSLEPARLKQEKQKLEL-----641  
 Db 511 WSSPHSILVNGATDYCSLSTSSSETANLNEHVGGQNLSEPIHOSPPARSSPELLPSG 570  
 Qy 642 EKQKEDAKRRQVQEDKQWLEHVQOEQPRPKPHEDRLKREDSVRKKEAER-----694

Db 571 VTDENEVTTAATEKVCSELDN-----NRHSEKEDPFVNDSSSLTGPAVDNLDPFQGS 622  
 QY 695 -----AKPEMODKOSRLFHHPHOEPK-----LATOAPWSTTEKGLPITSAQESV 738  
 Db 623 DPFVSGDPFKDDPFGKIDPFGGDPFKGSDPFASDCFFRQSTDPFATSSTDPSAANNSSI 682  
 QY 739 KVVYRRLYPFESRSHDEITIQPGDIVMDESOT-----GEGWLGC-----780  
 Db 683 TSV-----ETLKHNDPAPGCTVVAASDASATDPFASVFGNESFCGGFADPSTLSKVN 734  
 QY 781 ---ELKGTGMPFAN-----YAEKIPENEVPTPAKPVTDLTSAAPKLALRETPALP 830  
 Db 735 NEDPFRSATSSSVSNVVTIKNVFEETSVKSEDEPPALP-----PKIGTPTRPCPLP 785  
 QY 831 -----VTSSEP-----STTPNNMADFSSTWSSSNKPEPTDNWDT 865  
 Db 786 PGKRSINKLSDPDPFKLNDPFGPPGNDSPKEXDPEMFCDPFTSATTTTNKEADPSNFAN 845  
 QY 866 WAAQPS 871  
 Db 846 FSAVPS 851

RESULT 5  
 US-08-480-145-2  
 ; Sequence 2, Application US/08480145  
 ; Patent No. 5717067  
 ; GENERAL INFORMATION:  
 ; APPLICANT: DiFiore, Pier P  
 ; APPLICANT: Fazioli, Francesca  
 ; TITLE OF INVENTION: A Substrate for the Epidermal Growth  
 ; TITLE OF INVENTION: Factor Receptor Kinase  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Knobbe, Martens, Olson & Bear  
 ; STREET: 620 Newport Center Drive, Sixteenth Floor  
 ; CITY: Newport Beach  
 ; STATE: California  
 ; COUNTRY: United States of America  
 ; ZIP: 92660  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/480,145  
 ; FILING DATE: 07-JUN-1995  
 ; CLASSIFICATION: 530  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/095,737  
 ; FILING DATE: 22-JUL-1993  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Israelson, Ned A  
 ; REGISTRATION NUMBER: 29,655  
 ; REFERENCE/DOCKET NUMBER: NIH060.001A  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (619) 235-8550  
 ; TELEFAX: (619) 235-0176  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 896 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-480-145-2

Query Match 7.8%; Score 486.5; DB 1; Length 896;  
 Best Local Similarity 20.5%; Pred. No. 2,9e-26;  
 Matches 198; Conservative 150; Mismatches 273; Indels 345; Gaps 29;  
 QY 15 WAITVEERAKHQOQLSLKPIAGFITGDQARNPFQSGILPQVLAQIOWALADNMNDGMD 74

Db 122 WAVKPEDAKXDALPDSLSPVNGFSGDKVAPVLNLSKLPVLDLGRVWELSDIDHGMLD 181  
 QY 75 QVEFSIANKLILKILKQIGQLPSTLPVWKQOPVAISSAPAFGIGIASMPPLTAVAPVM 134  
 Db 182 RDEFVAMFVLVYCALE-----KEVPVM 203  
 QY 135 GSIPVWGSPPVSSVPPAAVPPLANGAPVIGPLPAFAHAPATWPKSSSFRSGSQSL 194  
 Db 204 -----SLPPALVP-----SKR 215  
 QY 195 NTKQAKQSPVVASAPPAEWAAPQSSRLKXRLQFNSHDKTMSGHLTGPOARTIMQSSL 254  
 Db 216 KT-----WVVSAPAKAKYDEIFLTKDMDQFVSGLEVEIFLTKGL 257  
 QY 255 POAOLASIWNLSDIDQGLTAEFFILAMHLIDVAM--SGQPLPVLPPEYIPPSFRVRS 313  
 Db 258 PSTLNAHIMSLCTDKGCKSLKQDFALAFHLISQKLTAKGIDPPHVLTPPEMIPPS--DRA 314  
 QY 314 GSGMSVSSSSVDQRLPEEPSSDEQOPEKKLPVTFEDKKRENFERGSELEKRRQALLE 373  
 Db 315 SLQKNIIIGSSPV-----ADFGAIK-----ELDTLNEIVD 344  
 QY 374 QQRKEQERLADQERAEQERKERERQEOBAKQOLELEKQLEKQERLEKQERREKTEIRR 433  
 Db 345 LQREK-----NNVEDDLKEK-----DTIKORISEVO--DLQDEVORE 380  
 QY 434 EAAKRELEQEROLEWERNRROELNQRNKEQEGTVLAKARKTLEFFELALNKKQOLEG 493  
 Db 381 NTNQIKQAOQKQ-----QOVELLDE-----LDQKQOLE 410  
 QY 494 KLQDIRCLATQROELISTNKSRELRJAETTHLOOQLOESQOMGRILIPKQILSDQLK 553  
 Db 411 QLKEVRKCAEAOILISLK-----AELTSQBSQISTYEELIAK-----AREELSR 456  
 QY 554 VQONSILRDSLLTLKRLAEKELARQOLREQDEVERETS-----KLQELIVFNNOJK 607  
 Db 457 LQOETAE-----LEESVESGKQALEPLQOHLDSQOELISSMOKLMEKDLNHNHNSQLN 510  
 QY 608 ELREIHS-----KQOLQKRSLEAARLKQEKQERKSLLE--641  
 Db 511 WCSSPHSILVNGATDYCSLSTSSSETANLNEHVGQSNLESEPIHQESPARSSPELLPSG 570  
 QY 642 EKQEDAKORVQERDKOMLHVQOEBOPRPRKHEDRLKREDSVKKAEER-----694  
 Db 571 VTDENEVTTAATEKVCSELDN-----NRHSEKEDPFVNDSSSLTGPAVDNLDPFQGS 622  
 QY 695 -----AKPEMODKOSRLFHHPHOEPK-----LATOAPWSTTEKGLPITSAQESV 738  
 Db 623 DPFVSGDPFKDDPFGKIDPFGGDPFKGSDPFASDCFFRQSTDPFATSSTDPSAANNSSI 682  
 QY 739 KVVYRRLYPFESRSHDEITIQPGDIVMDESOT-----GEGWLGC-----780  
 Db 683 TSV-----ETLKHNDPAPGCTVVAASDASATDPFASVFGNESFCGGFADPSTLSKVN 734  
 QY 781 ---ELKGTGMPFAN-----YAEKIPENEVPTPAKPVTDLTSAAPKLALRETPALP 830  
 Db 735 NEDPFRSATSSSVSNVVTIKNVFEETSVKSEDEPPALP-----PKIGTPTRPCPLP 785  
 QY 831 -----VTSSEP-----STTPNNMADFSSTWSSSNKPEPTDNWDT 865  
 Db 786 PGKRSINKLSDPDPFKLNDPFGPPGNDSPKEXDPEMFCDPFTSATTTTNKEADPSNFAN 845  
 QY 866 WAAQPS 871  
 Db 846 FSAVPS 851

RESULT 6  
 US-08-477-389-2  
 ; Sequence 2, Application US/08477389  
 ; Patent No. 5872219  
 ; GENERAL INFORMATION:

APPLICANT: DiFiore, Pier P  
 APPLICANT: Fazio, Francesca  
 TITLE OF INVENTION: A Substrate for the Epidermal Growth  
 FACTOR RECEPTOR KINASE  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Knobbe, Martens, Olson & Bear  
 STREET: 620 Newport Center Drive, Sixteenth Floor  
 CITY: Newport Beach  
 STATE: California  
 COUNTRY: United States of America  
 ZIP: 92660  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/477,389  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 530  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/095,737  
 FILING DATE: 22-JUL-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Israelsen, Ned A  
 REGISTRATION NUMBER: 29,655  
 REFERENCE/DOCKET NUMBER: NIH060.001A  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (619) 235-8550  
 TELEFAX: (619) 235-0176  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 896 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-477-389-2

Query Match 7.8%; Score 486.5; DB 2; Length 896;  
 Best local similarity 20.5%; Pred. No. 2,9e-26;  
 Matches 198; Conservative 150; Mismatches 273; Indels 345; Gaps 29;

QY 15 WAIVERRAKHQOFLSKPIAGITGPOARNFFOSGLPQPVLAQIMALAMNNDGRMD 74  
 DB 122 WAVEPBDKADAI FDSISPVNGFLSGDKKPVLLNSKLPVLDIGRWELSDIDHMDLD 181  
 QY 75 QVEFSIAMKLTKLKGQYLPSTLPPVKKQPVAISSAPAFGIGIASMPPLTVAVPM 134  
 DB 182 RDEFVAMFLVYCALE-----KEFVPM 203  
 QY 135 GSIPVGMSPVLSSVPPAANVPPLANGAPVYIOPAFAPHAATWPKSSSFRSGPSQL 194  
 DB 204 -----SLPPLALVP-----SKR 215  
 QY 195 NTKLQKAGSFVAPAPPAEWAPOSRLKYROLFNSHDKTMSGHLTGPARILLMOSL 254  
 DB 216 KT-----WVSPAERAKYDRIFLKTDKMDGVSLGLEVREIFLKTGL 257  
 QY 255 POAOLASIMNLSIDODGKLTAEFTILAMHLIDVAM-SGQPLPVLPPEYTPPSFRFRYS 313  
 DB 258 PSTLAIHWSLCDTKDCGKLSKQDFALAFHLISQKLIGIDPPHVLTPMEMPPS---DRA 314  
 QY 314 GSGMVISSSSDQRLPPEPSSSEDOPEKKLPVTEEDKKRNFERSGVELEKRRQALLE 373  
 DB 315 SLQKNITGSSPV-----ADFSATK-----ELDTLANEIVD 344  
 QY 374 OORKEORLAQLERAEREREROEAKROLELEKOLEKORLELORREERKEIERR 433  
 DB 345 LQREK-----NNVEQDLKEKE-----DTIKQRTSEVO---DLQDEVQRE 380  
 QY 434 EAKKELEOROLEWRNRROLNORNKKEQGVVLAARRKTELEFEALNDKKHOLEG 493

DB 381 NTNLQKLAQOKO-----QVLELDE-----LDEOKQALEE 410  
 QY 494 KLQDIRCLATOREROESTNKSRELRIATITHLQOQLOESQOMGRILPEKQILSDQKQ 553  
 DB 411 QKVEKRCCKAEBAQLISLK-----AELTSQESQISTEEELAK-----ARELSR 456  
 QY 554 VQONSLHRDLSLLTKRALAEKELARQQLREQLDEVERETRS-----KLOEIVFNQK 607  
 DB 457 LQETPAE-----LEESVESGKAQLEPLQOHLQDSQOQISMMQKMLMEMDLNHNQNLN 510  
 QY 608 ELREIHS-----KQLOKQSLAEARLKQKQERKSUEL-----641  
 DB 511 WCSSPHSILVNGATDYCSLSTSSSETANLNEHVGSNLSBPPIHOSPPARSSPELLPSG 570  
 QY 642 EKQXEDAQRVQERDKQMLEHVQOEOPRPKPHEEDRLKREDSVRKKEAER-----694  
 DB 571 VTDEHEVTAATEKVCSSLN-----NHSKEDPFPVNDSSLGVPADTVLDFQOS 622  
 QY 695 -----AKPEWQDKQSLFHPHOEPK-----LATQAPMSTTEKGLTISAGESV 738  
 DB 623 DPFVGSDFPKDDPFKIDPFQCDPFKSGDPPASDCFPQSTDPATSTDPFSANNSI 682  
 QY 739 KVVYRALYPPESRSHDITTOPGDIVWVDESOT-----GEGWLGQ-----780  
 DB 683 TSV-----ETLKHNDPPAPGGIVVAASDAITDPASVFGNSFGGCFADFTLSKVN 734  
 QY 781 ---ELKGTGWFFPAN-----YAEKIPENEVPTPAKEVTDLTSAAPAKLARETPALP 830  
 DB 735 NDDPPRSATSSSVSNVVTQKVFEETSVKSEDEPPALP-----PKIGTPTRPCPLP 785  
 QY 831 -----VTSESP-----STPPNMADFSSTWPMSSSNKPKETDMDT 865  
 DB 786 PGRKSNKLDSPPPKLNDPPQPPGNDSPKXKPEMFCDFTSATTTNKADPSNPNAN 845  
 QY 866 MAAPS 871  
 DB 846 FSAYPS 851

RESULT 7  
 US-08-095-737-4  
 Sequence 4, Application US/08095737  
 Patent No. 5487979  
 GENERAL INFORMATION:  
 APPLICANT: DiFiore, Pier P  
 APPLICANT: Fazio, Francesca  
 TITLE OF INVENTION: A Substrate for the Epidermal Growth  
 FACTOR RECEPTOR KINASE  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Knobbe, Martens, Olson & Bear  
 STREET: 620 Newport Center Drive, Sixteenth Floor  
 CITY: Newport Beach  
 STATE: California  
 COUNTRY: United States of America  
 ZIP: 92660  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/095,737  
 FILING DATE: 19930722  
 CLASSIFICATION: 530  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Israelsen, Ned A  
 REGISTRATION NUMBER: 29,655  
 REFERENCE/DOCKET NUMBER: NIH060.001A  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (619) 235-8550  
 TELEFAX: (619) 235-0176  
 INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 897 amino acids  
 TYPE: amino acid  
 TOPOLOGY: 1 linear  
 MOLECULE TYPE: protein  
 US-08-095-737-4

Query Match 7.8%; Score 486; DB 1; Length 897;  
 Best Local Similarity 21.3%; Pred. No. 3.2e-26;  
 Matches 205; Conservative 138; Mismatches 298; Indels 322; Gaps 31;

```

15 WATTEERAKHDQQLSLKPIAGFITGDQARNFFQSGLPQVLAQIMALAMNNDGRND 74
122 WAKSDEKAKYDAIFPSLSVDGSLGDKVKVPLNSKLPVELILGRVWELSDIDHDKLD 181
75 QVEFSIAMKLKIKLQGYQLPSTLPVPMKQPVVAISSAPAFIGIGIASMPPLTAAVAPVM 134
182 RDEFVAVMFLVYCALE-----KEPVPM 203
135 GSIPVGMSPPLVSSVPPAAPPPLANGAPPVIOPLPAFAHPAATWPKSSSFRSGPGSOL 194
204 -----SLPPLVPP-----SKR 215
195 NTKLQAKQSFVVASAPPAEMAVPOSSRLKYROLFNSHDKTMSGHLTGPQARTILMOSSL 254
216 KT-----WVSPAEKAKYDEIFLKTDKMDGYVSGLEVETFLKTGL 257
255 PQALASINNSLSDIDQDKLTAEEFLAMHLIDVAN-SSQPLPVLPPRYIPPSFRVRS 313
258 PSALLHINSLCCTKCGKSLKQFALAPHLINOKLIKIDIPHSITPMPIPPS--DRS 314
314 GSGMSVSSSDYQRLPEPPSSDEQPEKLPVTFEDKKRENFRGSLTELEKRRQALLE 373
315 SLQKNTGSSPV-----ADPSAIK-----ELDTLNNEIVD 344
374 QQRKEQERLAQLERAEQERERQERQAEKROLLEKOLEQERLEQERERKEIERR 433
345 LOREK-----NNVEQDLKEK-----DTVKQRTSEVQ--DLQDEVQRE 380
434 EAKKRELEROROLEMERNRNROELNORNKEGEGTVLTKARKKTLFELEALNDKHOLEG 493
381 SINLQKLOKQKQ-----OVQELGK-----IDEQAQOLEE 410
494 KLQDIRCLATOROLESTNKSRELRIAEITHLQOOLQESQOMLGRLLPEKQILSDQKQ 553
411 QLOEVKAKCAEBAQILSLK-----AEITQESQISSVEEELK-----AAEELSR 456
554 VQONSILHRSLTLTKRALEAKELARQQLREQLDEVERE-----TRSKIQEIDVFNQKLE 608
457 LQOETQK-----LEESVESGKAOLEPLOQHLQESQOEISSQMRLKEMKDETENNQ--- 507
609 LREIHSKQOLQKQ-----RSLERARKQKQERKSLLEKQEKEDQORVQ 653
508 -----SNSSSPQSVLVNATDYCSLSTSSSTANFNENABEONNLESPHIOESVASS 562
654 ER-----DKOMLEHVQOEOPRPKPKHEEDRLKREDSVRKKA-----EERAKPEMDK 702
563 PEIAPSDVLTDESEAVTVAGNEKVTPR--FDDDKHSKEDPEFVSSSLTDVADTNILDF 620
703 GSRLLF-----HPHOEPALQAPQWSTTEKPELILTSQESKVVYY 743
621 QSDPFGVSGPFDKDPFGKIDPFPGSDPFKSDCFPKQJSTDPPTTSTDPSPASNN 680
744 RALYPFESKSHDEITIPQDIIWVDESQGTBP--GWLGGELKKGKGMFPAYAEKIPENE 801
681 SNTSVEITWKHNDPPAPGGTVVAAASDASATDPFASVFNESFG--DGFADFSLKSVNND 739
802 VFTP-----AKPYTDLTSA-----PA--PKALAEPAPLPVTSSEPTTPNN 842
740 AKNPTISSTSVTLIAKPLEETASKSEDVPPALPPKGTPTRPCPP-----PGKRPIN 794
843 WADFEST-----WSSSNKEKETDNMDTMAAQPSLTVPSAGQLRQSAFTPTATQSS 895

```

Db 795 KLDSSDPLKINDPFPQPPGNDSPKXKDPDMFC-----DPFTSTTNKE 838  
 QY 896 PSP 898  
 Db 839 ADP 841

## RESULT 8

US-08-480-145-4  
 Sequence 4, Application US/08480145  
 Patent No. 571067

## GENERAL INFORMATION:

APPLICANT: DiFiore, Pier P  
 APPLICANT: Fazio, Francesca  
 TITLE OF INVENTION: A Substrate for the Epidermal Growth  
 TITLE OF INVENTION: Factor Receptor Kinase  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Knobbe, Martens, Olson & Bear  
 STREET: 620 Newport Center Drive, Sixteenth Floor  
 CITY: Newport Beach  
 STATE: California  
 COUNTRY: United States of America  
 ZIP: 92660

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/480,145  
 FILING DATE: 07-JUN-1995  
 CLASSIFICATION: 530  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: US 08/095,737  
 FILING DATE: 22-JUL-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Israel, Ned A  
 REGISTRATION NUMBER: 29,655  
 REFERENCE/DOCKET NUMBER: NIH060.001A  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (619) 235-8550  
 TELEFAX: (619) 235-0176  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 897 amino acids  
 TYPE: amino acid  
 TOPOLOGY: 1 linear  
 MOLECULE TYPE: protein  
 US-08-480-145-4

Query Match 7.8%; Score 486; DB 1; Length 897;  
 Best Local Similarity 21.3%; Pred. No. 3.2e-26;  
 Matches 205; Conservative 138; Mismatches 298; Indels 322; Gaps 31;

```

15 WATTEERAKHDQQLSLKPIAGFITGDQARNFFQSGLPQVLAQIMALAMNNDGRND 74
122 WAKSDEKAKYDAIFPSLSVDGSLGDKVKVPLNSKLPVELILGRVWELSDIDHDKLD 181
75 QVEFSIAMKLKIKLQGYQLPSTLPVPMKQPVVAISSAPAFIGIGIASMPPLTAAVAPVM 134
182 RDEFVAVMFLVYCALE-----KEPVPM 203
135 GSIPVGMSPPLVSSVPPAAPPPLANGAPPVIOPLPAFAHPAATWPKSSSFRSGPGSOL 194
204 -----SLPPLVPP-----SKR 215
195 NTKLQAKQSFVVASAPPAEMAVPOSSRLKYROLFNSHDKTMSGHLTGPQARTILMOSSL 254
216 KT-----WVSPAEKAKYDEIFLKTDKMDGYVSGLEVETFLKTGL 257
255 PQALASINNSLSDIDQDKLTAEEFLAMHLIDVAN-SSQPLPVLPPRYIPPSFRVRS 313

```

```

Db 258 PSALLAHISLCTKCGCKLSKDQFALFHLINOKLIKIDPHSHITPMPPS---DRS 314
Qy 314 GSGMSVSSSVYDQRLPEPSSSEDEQPEKKLVTFEDKRENFERSGVELEKRRQALLE 373
Db 315 SLQKNITGSSPV-----ADPSAIK-----ELDTLNNEIVD 344
Qy 374 QQRKEORLAQLERAEQERERQOEAKRQLELEKQLEKRELERQEEERKEIER 433
Db 345 LQREK-----NNVEQDLYKEKE-----DTVKQRTSEVQ---DLQDEVORE 380
Qy 434 EAAKRELERQRLERWERNRQELLNQRNKEQEGTVVLKARKKLEFELBALNDKKHQLEG 493
Db 381 SINQKLOAQOKQ-----QVQELLGE-----LDEQKQLEB 410
Qy 494 KLQDRCRLATQROEISTNKSRELRIAEITHLQOOLQESQOMLGRILPEKQILSDQLOK 553
Db 411 QLOEVRKKCAEBAQILSSIK-----AETTSQESQISSYEELIK-----AREELSR 456
Qy 554 VQONSILHRDSLTLTKRALEAKELARQQLREQLDEVERE-----TRSKLOEIDVFNQOLKE 608
Db 457 LQOETAQ-----LEESVESGKAQLEPLQOHLQESQOELISSMQRLMKDLDTNNQ--- 507
Qy 609 LREIHSKQLOLOKQ-----RSLBAARLKQKEQERKSLLELEKQEDAQRRVQ 653
Db 508 -----SNWSSSPQSVLVNGATDYCSLSTSSSETANFNENAGQNNLESEPTHQESSVRS 562
Qy 654 ER-----DKQMLEHYQOEOPRPKRPHEDRLKREDSYRKKEA-----EERAKPEMDK 702
Db 563 PELAPSDVYDESAVTVAGNEKVTPR--FDQDKHSKEEDPFNVESSSLDVAADTNLDF 620
Qy 703 QSRILF-----HPHQEPKATQAPWSTTEKGLTISAQESVKVYVY 743
Db 621 QSDPFGVQSDPFKDDPFKIDPFQGDPPKSPDPAQSCFFKQSTDPSTTSSTDPFASASN 680
Qy 744 RALYPRESRHDELTTPQDQIVWVDESQTEP--GMLGGLKQKTMFPANVAKIPEME 801
Db 681 SSNTSVETWGHNDPFAFGCTVVAASDASATDPPASVFGNSFG--DGFADFTSLSKVNNED 739
Qy 802 VPRP-----AKPVTDLTSA-----PA--PKLALRETPAPLPVTSSEPTPPNN 842
Db 740 ANFPPTISSSTSVTIKAPMLEETASKSEDPVPAKPRKVGTPTRPCPP-----PGKRPIN 794
Qy 843 WADFSST-----WPSSSNEKETDNDMTMAQPSLTVPSAQQLRQSAFTPATAGSS 895
Db 795 KLSQSDPLKINDPFPQRPNGDSPKEXPDMFC-----DPTSTSTTNNK 838
Qy 896 PSP 898
Db 839 ADP 841

```

```

RESULT 9
US-08-477-389-4
; Sequence 4, Application US/08477389
; Patent No. 5872219
; GENERAL INFORMATION:
; APPLICANT: DiFiore, Pier P
; TITLE OF INVENTION: A Substrate for the Epidermal Growth
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: United States of America
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,389
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/095,737
; FILING DATE: 22-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelson, Ned A
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH060,001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 897 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-477-389-4

```

```

Query Match 7.8%; Score 486; DB 2; Length 897;
Best local similarity 21.3%; Pred. No. 3,2e-26;
Matches 205; Conservative 138; Mismatches 298; Indels 322; Gaps 31;

```

```

Qy 15 WAITVEERAKHDOFLSKPIAGITGDQANRFPQSGLPQPVLAQIWMALAMNDRMD 74
Db 122 WAVSEDAKAYDALFDSLSVDFGSLGKVKVPLVNSKLPVEILIGRWELSDIDHCKLD 181
Qy 75 QVERSIAMKLIKLOQYQLPSTLPVWKQOPVAISSAPAGIGIASMPPLTAVAVPM 134
Db 182 RDEFVAMFLVYCALE-----KEFVPM 203
Qy 135 GSIPVVGSPPLVSSVPPAAPPPLANGAPVITQPLPAFHPAATWPKSSSPRSQSGOL 194
Db 204 -----SLPPLVLP-----SKR 215
Qy 195 NTKLOKQSPVNASAPPAEWAPOSSRLKYROLFNHDKTMSGHTGPQARTILMOSSL 254
Db 216 KT-----WVSGPAEKAKYDELFLTKDKMDGYVGLERETFLTKGL 257
Qy 255 POAQLASIMNSDIDQDKLTAEFFILAMHLIDVAM--SGQPLPVLPPEVTPSPFRVRS 313
Db 258 PSALLAHISLCTKCGCKLSKDQFALFHLINOKLIKIDPHSHITPMPPS---DRS 314
Qy 314 GSGMSVSSSVYDQRLPEPSSSEDEQPEKKLVTFEDKRENFERSGVELEKRRQALLE 373
Db 315 SLQKNITGSSPV-----ADPSAIK-----ELDTLNNEIVD 344
Qy 374 QQRKEORLAQLERAEQERERQOEAKRQLELEKQLEKRELERQEEERKEIER 433
Db 345 LQREK-----NNVEQDLYKEKE-----DTVKQRTSEVQ---DLQDEVORE 380
Qy 434 EAAKRELERQRLERWERNRQELLNQRNKEQEGTVVLKARKKLEFELBALNDKKHQLEG 493
Db 381 SINQKLOAQOKQ-----QVQELLGE-----LDEQKQLEB 410
Qy 494 KLQDRCRLATQROEISTNKSRELRIAEITHLQOOLQESQOMLGRILPEKQILSDQLOK 553
Db 411 QLOEVRKKCAEBAQILSSIK-----AETTSQESQISSYEELIK-----AREELSR 456
Qy 554 VQONSILHRDSLTLTKRALEAKELARQQLREQLDEVERE-----TRSKLOEIDVFNQOLKE 608
Db 457 LQOETAQ-----LEESVESGKAQLEPLQOHLQESQOELISSMQRLMKDLDTNNQ--- 507
Qy 609 LREIHSKQLOLOKQ-----RSLBAARLKQKEQERKSLLELEKQEDAQRRVQ 653
Db 508 -----SNWSSSPQSVLVNGATDYCSLSTSSSETANFNENAGQNNLESEPTHQESSVRS 562
Qy 654 ER-----DKQMLEHYQOEOPRPKRPHEDRLKREDSYRKKEA-----EERAKPEMDK 702
Db 563 PELAPSDVYDESAVTVAGNEKVTPR--FDQDKHSKEEDPFNVESSSLDVAADTNLDF 620

```







REFERENCE/DOCKET NUMBER: NIH054.001A  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (714) 760-0404  
 TELEFAX: (714) 760-9502  
 INFORMATION FOR SEQ ID NO: 94:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1898 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-644-94

Query Match 6.1%; Score 381; DB 2; Length 1898;  
 Best Local Similarity 27.0%; Pred. No. 2.7e-18;  
 Matches 136; Conservative 117; Mismatches 154; Indels 96; Gaps 21;

326 DQRLPEPSEDEQOPEKLPVTFEDKKRNPFERSVETLEKRRQALLLEQO--RKEOE--R 381  
 348 EQQLRR--QEERRRQQLRREGEEERRRQQLRREGEEERRRQQLRREGQQLRRQQLR 404  
 382 LAQLERAEQERKE--RQOEBAK--OLELEKOLEKORLEERQEEER----- 426  
 405 EQQLRRQQLRREGQQLRREGQQLRREGQQLRREGQQLRREGQQLRREGQQLRREGQQLR 464  
 427 ---RKEIEREAKRLEEROROLEWENRRROELNQNKEOGTVVLKARKTLFELEA 483  
 465 RLKREOEERDMLKKEETERH--EOERRKOQ--KDOEE--RRRWLKLEEE- 513  
 484 LNDKHOLESKLDIRICRLATORLESTNKSRLIAETLHLOQLQESQ-----QML- 537  
 514 -ERREQOEERQQLR-REOEERREQLKQEEERLQQLRSEQQLRRQOEERLQLLK 570  
 538 -----GRLLPEKQILSDQ---KOVQNSLRHSDSLTLTKRLAKELARQO 580  
 571 REBEKRLQERRRQQLRREGQERRDQQLKEEERQQLRREGQERRLKEEVEERLE 630  
 581 LREOLDE-VERETSKLQELIDVENNQLKELR---EHSKOQLKQPSL---EARRLKOK 632  
 631 QEBERDRRLKREBEERREHLLKSEQEEERRHQQLRREGQERRLKREBEERLEQ 690  
 633 ---EOERSLLEL-EKQEDAQRRVOERDKQMLEHVQOEQPRPK----- 673  
 691 LKREHEERREOELEAEEOQARERIKSRIPKQWQLESFADARQSKVLLEAPQARABA 750  
 674 PHEEDRLKRDSPYKKAERAKPEMODKOSPLFPHQOEBAKATAPMSTTEKGLTS 733  
 751 PQOEERKRRESLEQOEERRAHQOEERQDRDFTWQOAE-----EKSEGRQRLS 803  
 734 AOESEKVVYVYALYPPESRSHDE 756  
 804 APPPLRQERERQLRAERQOQREQ 826

RESULT 14  
 US-08-475-894-2  
 Sequence 2, Application US/08475894  
 Patent No. 5641748  
 GENERAL INFORMATION:  
 APPLICANT: Yen-Ming Hsu  
 TITLE OF INVENTION: THE CAIP-LIKE GENE FAMILY  
 NUMBER OF SEQUENCES: 6  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: LAHYVE & COCKFIELD  
 STREET: 60 State Street, Suite 510  
 CITY: Boston  
 STATE: Massachusetts  
 COUNTRY: USA  
 ZIP: 02109-1875  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Releasee #1.0, Version #1.25

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/475,894  
 FILING DATE:  
 CLASSIFICATION: 530  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Louis Myers  
 REGISTRATION NUMBER: 35,965  
 REFERENCE/DOCKET NUMBER: BGP-191  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 227-7400  
 TELEFAX: (617) 227-5941  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 553 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 FRAGMENT TYPE: internal  
 US-08-475-894-2

Query Match 5.3%; Score 333; DB 1; Length 553;  
 Best Local Similarity 27.4%; Pred. No. 1.2e-15;  
 Matches 93; Conservative 67; Mismatches 115; Indels 64; Gaps 15;

912 QALYPMARKDNHLPNKSQDVTVLLEQD--MMWGEVQOGKMPKSYKLSGPRKST 970  
 77 QDIYPLTRHDELTLISGELITNIRKEDGMMBGOJNGRRGLFPDVFREIKKMKKDP 136  
 971 SIDGPTSPASLRVAPAKPAIPGEETI-----AMYTESSBOGDLTF 1016  
 137 LTNKAP-EKP--LHEV--PSGNSLSSETLITRTKGERRRRRRCQVAFSLPQNDDEL 191  
 1017 QCGDVI-VYTKKGDWMTGYGDKSGVFPSPNYVA-----LKDSE 1054  
 192 KVGDIIEVVGVEGMEGVLNGKTGMPSPNFKELSGSDELGISQDEQLSKSSLETT 251  
 1055 GSGTAGTGSIGKKPEIAQVASYAATGPPEOL-----TLAPQGLILIRKKNPGMWGE 1108  
 252 GSESGGSS--STSGANGVATATAIQPKVKGVGCDLFXOKPIRLRRS-----IEVE 306  
 1109 LQARKKRQIG-WFPANYKLLSPGTSKITPTELKTAQVAPACQVIGMYDYRQNDDEL 1167  
 307 NDFLPEVEKTIKKLPAT---TATPDSSK--TEWDSRTSKSDYCKVI--PPYEQNDDEL 358  
 1168 AFSGQIINVLNKE--DPMWKGESGVGOVGLFPSPNYKL 1204  
 359 TIKSGDITVLINKDCIDVGMWGBLNGRGVFPDNYKL 397

RESULT 15  
 US-08-484-710-2  
 Sequence 2, Application US/08484710  
 Patent No. 5656438  
 GENERAL INFORMATION:  
 APPLICANT: Yen-Ming Hsu  
 TITLE OF INVENTION: THE CAIP-LIKE GENE FAMILY  
 NUMBER OF SEQUENCES: 6  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: LAHYVE & COCKFIELD  
 STREET: 60 State Street, Suite 510  
 CITY: Boston  
 STATE: Massachusetts  
 COUNTRY: USA  
 ZIP: 02109-1875  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Releasee #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/484,710  
 FILING DATE:  
 CLASSIFICATION: 530

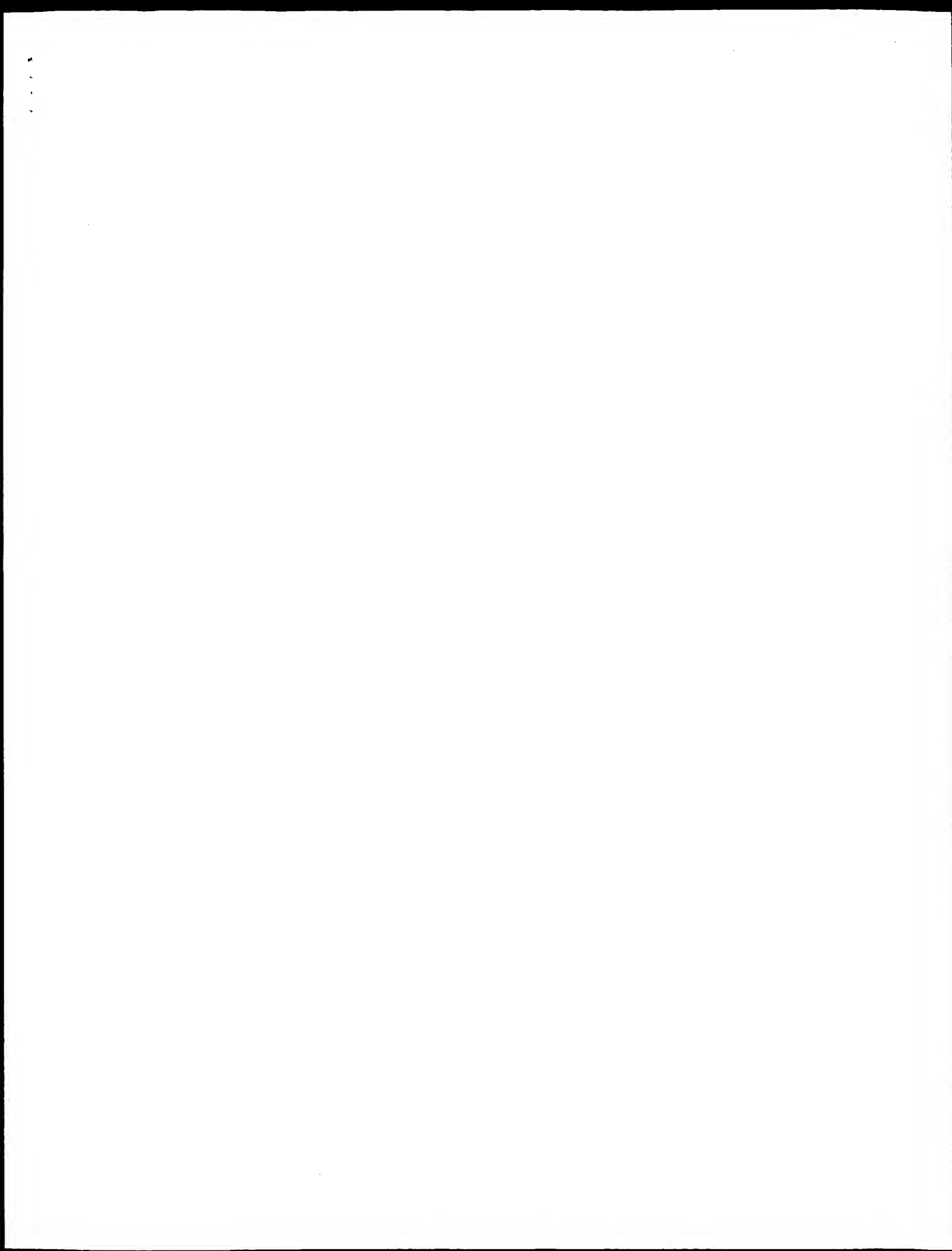
## ATTORNEY/AGENT INFORMATION:

NAME: Louis Myers  
REGISTRATION NUMBER: 35,965  
REFERENCE/DOCKET NUMBER: BGP-190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 553 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-484-710-2

Query Match 5.3%; Score 333; DB 1; Length 553;  
Best Local Similarity 27.4%; Pred. No. 12e-15;  
Matches 93; Conservative 67; Mismatches 115; Indels 64; Gaps 15;

QY 912 QALYPMWRAKKDNHLENKSDVITVLEQD-MWMEVQGGKMFPSYVYKLSGPYRKST 970  
Db 77 QDIYPLYLRHDDDELISVGEIITIRKEDGXWMEQINGRGLFPDNFVREIKKMKXDP 136  
QY 971 SIDTGPTESPASIKRYASPAKPAIPGEPI-----AMTYESESQDLTF 1016  
Db 137 LTNKAP-EKP-LHEV--PSGNSLSSETILRTNKRGERRRRCQVAFSYLPONDDLEL 191  
QY 1017 QOGDVI-VVTKDDGMWMTGVGDKSGVFPNSYVR-----LKDSB 1054  
Db 192 KVGDIIEVVEVEEGWEGVLNGKTGMFPNFIKELSGESDELGISODEQLSKSLRETT 251  
QY 1055 GSGTAGKTGSLGKKPEIAQVIASVAAATGPEQI-----TLAPQILIRKKNPGWMEGE 1108  
Db 252 GSESDGGDSS-STKSEGANGTAVATAIQPKYKVGVGFDIFKDKPIKLRPRS---IEVE 306  
QY 1109 LQARGKKRQIG-WFPANYVLLSPGTSKITPTLEPKTAVQVAVQVIGMTDYTAQNDDEL 1167  
Db 307 NDFLPVEKTIIGKKLPAT--TATPDSSK--TMDSRTSKSDYCKVI--FPYEAQNDDEL 358  
QY 1168 AFSKGOIINVLNKE--DPDMWKGEVSGOVGLFPSNYVKL 1204  
Db 359 TKEGDIIVTLINKDCIDVGMWEGELNGRGVFPDNFVKL 397

Search completed: March 14, 2003, 12:18:27  
Job time : 32 secs





Db	369	CCACTAATCTCTGCTCGTTTGGATGGAGACATGCCAATCTGTCATTCAATCAGCCA	428
Oy	634	CTCACAGCTGTGTGCTCTGTGCCAA-----TGGGCTCCATT	669
Db	429	TTCGCTCCAGTTGGACCTAATAGCAACACCCTGTCTTCTGTCTACTTTCAGGACACAGTATT	488
Oy	670	CCAAGTTTGGAAATGTCTCCACCCCTTATCTTGTCTCTCCCTCCAGACAGAGTGCCTCC	729
Db	489	CTCTCCCTTAATGATGCTGTCTCCCTGTAGTGTCTTGTATGTAATCTCACTTATTAACCAAT	548
Oy	730	CTGGCTAAACGGGGCTCTCCCGGTATACAGCTCTGTGCTCGTTTGGCATCTTGAGCC	789
Db	549	GGAACTGGCAGTCTCATTCACCCCTTATTCATTCTTATTTCTTCTTCAACATTGGCTTAT	608
Oy	790	ACATGGCCAAAGATCTTCTCTTCAGCAATCTGTCTCCAGGTCACATTAATAACATAAG	849
Db	609	GCATCATCTTACAGCTGATGATGGGAGATTGGTGCTAGTATCCAGAAAGCCAG	668
Oy	850	TTACGAAGGCAATCATTTGATGTGCCAGGCCCCCTCCAG-----	892
Db	669	TCTCTGATTTGATTTAGATCTTAGTACTCAATCTCTCAATCTGTCTCCCTTCAGGAAAC	728
Oy	893	-----CAGCAAAATGGGCTGTGCTCAGTCATCAAGGTCGAAATACAG	936
Db	729	TCACCTAAGACAGGGACCTCGAATGGGCACTTCTCAGCTTCAAGATTAAGATGG	788
Oy	937	CAGTATTCAACAGCGACGACAAACATATAGTGGACATTAACAGTCTCCAGGCAAGA	996
Db	789	CAAAAATTTAATGATCTTAGACAAAGGATAGGGATACCTCTCAGATTTTCAACTAGA	848
Oy	997	ACTATTCTCATGCAATCAAGTTTAAACCCAGGCTCAGCTGGCTTCAATATGAAATCTTCT	1056
Db	849	AATGCCCTTCTTCACTCAATATCTCTCAAACTCAGCTAGCTAATTTTGGACTCTGGCT	908
Oy	1057	GACATTGATCAAGATGCAAAACTACTGCAAGAAATTTATCTTACTATATGCACCTAAT	1116
Db	909	GACATCGATGTGACGCAAGTTGAAAGCTGAAGAAATTTATTTGTGGCGATGCACCTCACT	968
Oy	1117	GATGTTCATGTGTGTGCAGCCACTGCTGCGCCGCTCTGCTCTCGAATAACATCCCTCT	1176
Db	969	GACATGGCAAGTGTGGACAGCCACTACCTACCTAGCTTGTCTCCGAGTTGTCTCTCA	1028
Oy	1177	TCCTTCAGAGAGTTCCGCTCCGGCAGTGGAGTCCGTCATAGCTCTTTCTGTGTGAT	1236
Db	1029	TCCTTCAGAGG-----GGGAAACCAAGTTGATCTCTTTAAT	1064
Oy	1237	CAGAGGCTGCTGAGGAGCCGTGTACAGAGATGACACACACCTCAGAGAAACCTGCT	1286
Db	1065	GGAACTCTGCTTATATTCAGAAAACCAAGAG--AAGACCTCGAAGAAATTTGCA	1121
Oy	1297	GTGACATTGAAAGATAAGAAACGGGAAACTTTGACGAGGACAGTGTGGAGCTGGAGAG	1356
Db	1122	GTATCTTTTGGAGCAACCGAAAGCCAACTATGACGAGAAACATGGAGCTGGAGAG	1181
Oy	1357	CGCCGCCAAGCGCTCTTGGAGACAGACCGGAAAGACAGAGACGGTGTGCTCAGCTGAG	1416
Db	1182	CGAGCCAAAGTTGATGAGACAGCACAGACAGGGAAGCTGAACGCAAAACCCACAGAAAG	1241
Oy	1417	CGCCGCCAGGAGGAGAGAAAGACGGGAGCCGACAGACAGAGAGGCCCAAGCGCAGCTG	1476
Db	1242	AAGGAAAGTGGGAGCGGAAACAGAGAGAACTGTCAAGACCAAGAAATGGAAAGACACTG	1301
Oy	1477	GAGCTGGAGAAAGACGCTGGAGAAACAGACGGAGACTGTGAGCGCCAGCGAGAGAGAGAG	1536
Db	1302	GAGTTGAGAGAAACCTTTGGAGAAACAGAGAGCTGTGAAGACACCGGAGAGAAAGAGG	1361
Oy	1537	AGGAAAGAGATCGAGAGCGCCGAGGCGCAAAACGGGAACTGTGAAGGACACGCAACTT	1596
Db	1362	AGGAAAGAGATAGAAAGACAGAGGCGACGAAACAGAGACTTTGAAGACAAACGCCGTTTA	1421
Oy	1597	GAATGGAAACGGAACCGGAGACAGAACTCTTAATCAGAGGAAACAAAGACAGAGAGGC	1656

Db	1422	GAATGGGAAAAGACTTCCTCGTCCGACAGAGAGCTGCTCAGTCAGAAAACCAAGGACCAAGAGAC	1481
Qy	1657	ACCGTGGCTCTGAAGAGCAAGAGAGAGACTCTGAGATTGGATTGAGAAAGCTCTGATGAC	1716
Db	1482	ATTGTACAGGCTGACCTCCAGAAAAGAAAGTCTCCACTGGAACTGGAAACACTGATGATGA	1541
Qy	1717	AAAAAGCATCAGCTAGAGAGGAAAATTCCAGGATATCAGGTGTGCACTGGCAACCAAGG	1776
Db	1542	AAACATCAGCGAGATCTCAGGCAAGACTCAAGAATGTCCAATCAGAAAAGCAACAAAG	1601
Qy	1777	CAAGAAATTGAGACACGAACAAGTCTAGAGACTAAGAAATTCGAAATCAACCACTTA	1836
Db	1602	ACTGAGCTAGAAAGTTTGGATTAACAGCTGACCTGAAATTTATGAAATTCAAACACTT	1661
Qy	1837	CAGACGACGTTGCAGGAATCTCAGCAAAATCCTTGGAAAGCTTAATTCAGAGAAACAGTA	1896
Db	1662	CAACAGAGGCTTTAAGAAATTCAAAATTAAGCTTATCTATCTGGTCCCTGAGAGACGCTA	1721
Qy	1897	CTCAGTGACGAGTTAAACAAGTCCAGCGAGAACAGTTTSCATATAGAGACTCGCTTTTAC	1956
Db	1722	TTTAAACGAAAGAAATTTAAACATGCAAGCTCAGTACACACCTGATTCAGGGATCAGTTTA	1781
Qy	1957	CTCAAAAGAGCTTGGAAAGCAAGAGAGCTGCGCCGACAGACCTCGGAGAGCAGCTGAC	2016
Db	1782	CTTCATATAAAAGTTCATCAGAAAAGGAAGATTAATCCCAAGACTTAAGAACAAATTAAT	1841
Qy	2017	GAGTGGAGAGAGAGACCAGGTCAAAAGCTGACAGAGATTGATGTTTTCAACAACCACTG	2076
Db	1842	GCTCTGAAAAAGAACTGCATTAAGCTCTCAGAAATGGAATTCATTAAACATATAGCTG	1901
Qy	2077	AAGGACTGAGAGAGATACATAGCAAAACAGCAACTCCAAAGCAAGAGTCCCTGGAGCA	2136
Db	1902	AAGAACTCAGAGAAAGCTATATAACAGACAGTTAGCCCTTGAACAACTTATATAATC	1961
Qy	2137	GCGGCACTGAAGCAAAAGAGCAGGAGAGAGAAAGCTTGGATTAGAGACGAAA	2192
Db	1962	AAAGTGACAAATTTGAAGGAATTCGAAAGAAAAAGATTAGACAAAAA	2017
RESULT 2			
US-08-630-915A-193			
Sequence 193, Application US/08630915A			
Patent No. 6309820			
GENERAL INFORMATION:			
APPLICANT: SPARKS, Andrew B.			
APPLICANT: HOFFMAN, No. 6309820h			
APPLICANT: KAY, Brian K.			
APPLICANT: FOMKES, Dana M.			
APPLICANT: MCCONNELL, Stephen J.			
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL			
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND			
NUMBER OF SEQUENCES: 227			
CORRESPONDENCE ADDRESS:			
ADDRESSER: Pennie & Edmonds LLP			
STREET: 1155 Avenue of the Americas			
CITY: New York			
STATE: New York			
COUNTRY: USA			
ZIP: 10036-2711			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Floppy disk			
COMPUTER: IBM PC compatible			
OPERATING SYSTEM: PC-DOS/MS-DOS			
SOFTWARE: PatentIn Release #1.0, Version #1.30			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/08/630,915A			
FILING DATE: 03-APR-1996			
CLASSIFICATION: 356			
ATTORNEY/AGENT INFORMATION:			
NAME: Mistock, S. Leslie			
REGISTRATION NUMBER: 18,872			
REFERENCE/DOCKET NUMBER: 1101-174			

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 790-9090  
 TELEFAX: (212) 869-8864/9741  
 INFORMATION FOR SEQ ID NO: 193:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2873 bases  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: unknown  
 MOLECULE TYPE: DNA  
 US-08-630-915A-193

Query Match 6.6%; Score 333.2; DB 4; Length 2873;  
 Best Local Similarity 61.0%; Pred. No. 1e-76;  
 Matches 633; Conservative 0; Mismatches 363; Indels 42; Gaps 4;

QY 2867 CTCTCTGACCGTACCTAGTGTGGCCGATTACGGCAGAGATCAGCTTTACCCCAAGCA 2926  
 Db 541 CTTTTCAAACCTAAGCTTAATATACATCATGCGCAAAAAAATCAGCTTCACTCGAAGCTG 600  
 QY 2927 CAGCCACTGCTCTCTCCCATCTCCCTCTGGCCGAGGGTGAAGGTGGAAGGGCTAC 2986  
 Db 601 TGTCCCTCTG---ATCTGTACCTAATTCATGACGAGCAAGGTGTAAGAACTTAA 657  
 QY 2987 AAGCGCAAGCCGTATCCCTGGAGAGCCAAAAAGACACCACTTAATTTTAACAAAA 3046  
 Db 658 AAGCAGAGCCCTTGTCTCTGAGCTCAAGAAAGATTAACCACTTGAACCTTCAAAAC 717  
 QY 3047 GTGACGTATACCCGTTCTGGAACAGACATGTGTGTTGGAGAACTTCAAGATC 3106  
 Db 718 ATGACATTATTAATCTGTGTGAGAGCAGCAAAAAATGTGTGTTGGAGAGTGCATGAG 777  
 QY 3107 AGAAGGTTGTTCCCAAGTCTTACTGAAACCTATTTGAGGGCCCGCTAAGAAATCCA 3166  
 Db 778 GAAGAGATGTTTCCCAATCTTATGTCAAGATCTCTGGAGTGAATGAATGAAGCGG 837  
 QY 3167 CAAGCATGATGATGAGCCCTACTGAAAGTCTGTACTTAAGAGAGTGGCTTCCCGG 3226  
 Db 838 AA-----GAACCAAGCTTGTATGACGCTGTAATAAATAAACCCTACCTCGG 885  
 QY 3227 CCGCAGAGCCAGCATTCGCGAGAGAGTTTATTCGATGACATACAGAGTTCTG 3286  
 Db 886 CAGCTATTTCAGT-----TGGAGAGAAATATTGTCATTTTCCATATTCATGATG 939  
 QY 3287 AGCAAGAGATTTAACTTTAGCAAGAGGATGTGATTTGCTTACCAAGAAAGATGGT 3346  
 Db 940 AACCTGAGATTTGATTTCAAGAGAGTGAAGAAATATTGTTGACCCAGAAAGATGAG 999  
 QY 3347 ACTGTGAGCGGAAAGGTTGGGCGCAAGTCCGAGTCTTCCCTTCTAATATGTGAGC 3406  
 Db 1000 AGTGTGACAGGAAGTATGAGATAGAAATGGAATTTTCCATCAAACTATGTCAAC 1059  
 QY 3407 TTAAGATTCAAGAGGCTCTGAACTGCTGGGAAAAAGGAGATTGAGAAAAAAGCTG 3466  
 Db 1060 CAAGGATCAAGAGGTTTGGAGTCTGAGCAAGTCTGAGCATCAATTAATAAAGCCTG 1119  
 QY 3467 AAATTGCCAGATTATTCCTTCTTAAGCTGCTACTGCTCCGGAACAATCACTGGCTC 3526  
 Db 1120 AGATTGCTCAGATTAATTCACATATGTTGCTTCTGTTCTGAAACAACTTAGCTTGG 1179  
 QY 3527 CTGGGCACTGATTTCTGATCCGAAAAAGAAAGCCAGTGTGATGTGTGAGAGAGAACTGC 3586  
 Db 1180 CAGAGCACTTATATTAATTTCTAAAGAAAAATACAGTGGGTGGGCAAGAGAGTTAC 1239  
 QY 3587 AAGCTGAGGAAAAAGCCAGATAGGTGTTTCCAGCAAAATATATCTCAACTTTAA 3646  
 Db 1240 AGGCGAGAGAAAAAGGAGACAGAAAGATGTTCTCTGCAAGTATGTTAACTTTGG 1299  
 QY 3647 GCGCCGGAACAGCAAAATCAACCCCACTAGCTACCAAGCCGCAATGACGACAGAG 3706  
 Db 1300 GTCCAAGTATGTAAGAGCCACCTGCTTTTCATC-----CTG 1338

QY 3707 TGTCCAGGTGATGGGATGTACATTTACACCCCGCCAGAACATGACATGACCTTCA 3766  
 Db 1339 TATGTCAAGTATGATGTATGATGACTATCAGCAAAATATGAAGATGACCTCACTTCT 1398  
 QY 3767 GCAAGGCCAGATCATCAAGTCTCTCAACAGAGAGAGCCGAGCTGTGAAAAAGAGAG 3826  
 Db 1399 CCAAGGACCACTATTAATTTATGAAACAAAGTATCTGTATGTTGGCAAGAGAGA 1458  
 QY 3827 TCAGTGGCAAGTTGGGCTCTTCCCATCAATTAATGTAAAGCTGACCAAGACATGAGC 3886  
 Db 1459 TCAAGGGGTGACTGTCTCTTCTTCAAACTACGTTAAGATGACGACAGACTGATC 1518  
 QY 3887 CCAGCCAGCAATGAATCA 3904  
 Db 1519 CAAGTCAACAGTGACCA 1536

RESULT 3  
 US-08-630-915A-39  
 Sequence 39, Application US/08630915A  
 Patent No. 6309820  
 GENERAL INFORMATION:  
 APPLICANT: SPARKS, Andrew B.  
 APPLICANT: HOFFMAN, No. 6309820h  
 APPLICANT: KAY, Brian K.  
 APPLICANT: FOWLES, Dana M.  
 APPLICANT: McCONNELL, Stephen J.  
 TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL  
 TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
 NUMBER OF SEQUENCES: 227  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: Pennie & Edmonds LLP  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10036-2711  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/630, 915A  
 FILING DATE: 03-APR-1996  
 CLASSIFICATION: 536  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Mistrock, S. Leslie  
 REGISTRATION NUMBER: 18,872  
 REFERENCE/DOCKET NUMBER: 1101-174  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 790-9090  
 TELEFAX: (212) 869-8864/9741  
 INFORMATION FOR SEQ ID NO: 39:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 747 bases  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA  
 US-08-630-915A-39  
 Query Match 5.1%; Score 258; DB 4; Length 747;  
 Best Local Similarity 63.9%; Pred. No. 1.8e-57;  
 Matches 418; Conservative 0; Mismatches 215; Indels 21; Gaps 1;

QY 3247 GGAAGAGATTATTTCCATGATACATACGAGATTCTGAGCAAGAGATTAACTTT 3306  
 Db 115 GGAAGAGATTATTTCCATGATACATACGAGATTCTGAGCAAGAGATTAACTTT 174  
 QY 3307 CAGCAAGGATGTGATTTGTGTTACCAAGAAAGATGACTGGTGAAGGAGAGGTG 3366

```

Db      175 ACAGAGGTGAAGAAATATGTGTACCCAGAGAGATGAGAGTGTGAGACGAGAAATATT 234
QY      3367 GGCACAAAGTCCGAGTCTTCCCTTCTAATATGTGAGGCTTAAAGTTGAGAGGGCTCT 3426
Db      235 GGAATATGAAGTGAATTTTTCATCAAACTATGTCAAAACCAAGATCAAGAGAGTTT 294
QY      3427 GGAATCTGTGGAAAACAGGAGTTTAGAAAAAACCTGAAAATTGCCAGTTATTGCT 3486
Db      295 GGGAGTGTCTAGCAAGTCTGGAGCATCAATTAATAAAACCTGAGATTGCTGATCACTTCA 354
QY      3487 TCCAGCTGTACTGTGTCGCCGAACAACCTACCTGTGCTCTGGAGAGCTGATTCGATC 3546
Db      355 GCATATGTGTTCTTGTGTTCTGAACAACCTTACCTTGACAGAGACAGTTAAATTAATT 414
QY      3547 CGGAAAAAGAACCCAGGTGATGTGTGGAAAGAGAACTGCAAGCTGAGGAAAAAGCCG 3606
Db      415 CTAAAGAAAATACAAAGTGGTGGTGGCAAGAGATTACAGGCCAGAGSAAAAAGCCGA 474
QY      3607 CAGATAGGTTGTTTCCAGCAAAATATGTCAAACTTCTAAGCCCGGACACGAAAAATC 3666
Db      475 CAGAAAGATGTTTCTGTCAGTCATGTTAACTTTGGGTCCAAGCATGMAAGAGCC 534
QY      3667 ACCCACTAGCTACCCAGACCGCAGTGCAGCCAGCATGTGTCAGGATCGGGATG 3726
Db      535 ACACCTGCTTTTATCTCT-----GTATGTACAGTGATTCCTATG 573
QY      3727 TACGATTACACCCGCAAGAACTATGACGAACTAGCTTCAGCAAGGCCAGATCATCAAC 3786
Db      574 TATGACTATCAGCAAAATATGAAGATGAGCTCAGTTTCTCAAGGACCACTATTAAAT 633
QY      3787 GTCTCTCAACAAGAGAGCCGCGAGTGTGSAAGAGAAAGTCTGTGGCAAGTTGGCTTC 3846
Db      634 GTTATGAACAAAGATGATCTCTGATTTGTCAGAGAGATCAACGGGTGACTGTCTC 693
QY      3847 TTCCCATTCATTTATGTAAGCTGACACAGCATGAGACCCGACGAGAATGA 3900
Db      694 TTTCCTTCAAACTAGCTTAGATGACGACGACTCAGATCCAAAGTCAACAGTGA 747

```

```

RESULT 4
US-09-404-879A-5
; Sequence 5, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-404-879A-5

```

```

Query Match      3.0%; Score 151.4; DB 4; Length 531;
Best Local Similarity 61.3%; Pred. No. 9,5e-30;
Matches 264; Conservative 0; Mismatches 161; Indels 6; Gaps 1;

```

```

QY      226 GGGCTGATCAGCAAGCTAATATAACCATGCTCACTTTCCACACTTTTGGT 285
Db      9 GGCTGAGACTGCAAGAAAGTCAAGATCATGATGCTCACTTTCCACACGATGAAT 68
QY      286 GGATAGCTGATGTCTGGCCATAACTGTGAGAGAAAGGCGCACAGATACAGAGATT 345
Db      69 GGAGGGCCAAATATGTGGCTATTACATCTGAAGAAAGTACTAAGCATGATAAAGATT 128

```

```

QY      346 CTTAGCTGAAGCCGATAGCCGATTTTATTACTGTGATCAAGGAGAACTTTTTC 405
Db      129 GATAACCTCAAACTTTCAGAGAGTTTACATAACAGTATCAAGCCCGATCTTTTCTTA 188
QY      406 CAATGTGGTTACTCAAGCTCTCTTAGACAAATATGGCGCTTAGGAGCATGATAAC 465
Db      189 CAGTAAAGTCTGCGGCCCGCTTTTAGCTGAATATAGGGCTTATCATGATCTGAACAG 248
QY      466 GATGAGAGATGATCAAGTGAATTTTCCATAGCCATGAAAGCTTATCAAGTGAAGCTA 525
Db      249 GATGGAAGATGAGACAGCAAGAGTTCTTATAGCTATGAACTCATGATTAAGTTG 308
QY      526 CAAGATATCAGTCTCCCTCCACACTTCCCTGTGATGAAACAGCAACAGTGGCTATT 585
Db      309 CAGGGCCAAACAGCTGCTGTAGTCTCTCTCTCATGAAACAAACCC-----CTATG 362
QY      586 TCAGTGACAGCAGATTTTGTATAGAGAGATTCTAGATGCCACATCAGACTGTT 645
Db      363 TTCTCTCCTAATCTCTGCTGTTTGGAGTGGAAAGCATGCCAAATGTCATTCAT 422
QY      646 GCTCCTGTGCC 656
Db      423 CAGCATTTGCC 433

```

```

RESULT 5
US-09-404-879A-60
; Sequence 60, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 60
; LENGTH: 480
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-404-879A-60

```

```

Query Match      2.8%; Score 144; DB 4; Length 480;
Best Local Similarity 62.1%; Pred. No. 7,6e-28;
Matches 247; Conservative 0; Mismatches 145; Indels 6; Gaps 1;

```

```

QY      259 ATGGCTAGTTTCCACACTTTCGGTGTAGCTGTGATGTCTGGGCATTAATCTTGAG 318
Db      14 ATGGCTAGTTTCCACAGGATGAATGAGGGCCAAATATGTGGCTATTACATCTGAA 73
QY      319 GAAAGGCCAAGCATGACCGCAGTCTCTTACGCTGAAGCCGATAGCGGATTTATTACT 378
Db      74 GAAGTACTAAGCATGATTAACAGTTTGATTAACCTTCAAGACTCAGAGATTAATAA 133
QY      379 GGTATGAAGGAGAACTTTTTCATCTGAGTTACTTCAAGCTGTCTTACACAA 438
Db      134 GGTGATCAAGCCCGTACTTTTCTTACAGTCAAGTCTGCGGCCCGGTTTAGCTGAA 193
QY      439 ATATGGCGCTAGCGGCAATGATTAAGATGAGAGATGATTAAGTGAATTTTCCATA 498
Db      194 ATATGGCTTATCAATCTGAAACAGATGGAGATGACACGAAAGATTTCTTATA 253
QY      499 GCGATGAAGCTTAACTGAAGCTTACAAAGATATACAGTCCCTCCACACTTCCCT 558
Db      254 GCTATGAACCTCAATCAAGTTAAAGTTGACAGGCGCAACACTGCTGTATCTCTCTCT 313
QY      559 GTATGAAACAGAACACAGTGGCTATTTCCAGTGACCGCAGCATTTGTATGAGAGATT 618
Db      314 ATCATGAACAAACCC-----CTATGTTCTCTCACTAATCTCTGCTGTTTGGAGT 367

```



QY 619 GCTAGCATGCCACCACTCAAGCTGTGCTCTCTGCC 656  
 Db 368 GGAAGCATGCCCAATCTGTCTCATTCATCAAGCATTCGCC 405

RESULT 6  
 US-08-728-323A-1  
 ; Sequence 1, Application US/08728323A  
 ; Patent No. 5948676

## GENERAL INFORMATION:

APPLICANT: Chang, Yuan  
 APPLICANT: Bohenzky, Roy A.  
 APPLICANT: Russo, James J.  
 APPLICANT: Edelman, Isidore S.  
 APPLICANT: Moore, Patrick S.  
 TITLE OF INVENTION: Immediate Early Protein From Kaposi's  
 TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA  
 NUMBER OF SEQUENCES: 21  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Cooper & Dunham LLP  
 STREET: 1185 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: U.S.A.  
 ZIP: 10036

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/728,323A

## FILING DATE:

CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: White, John P.

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212-278-0400

TELEFAX: 212-391-0525

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 3489 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 1..3489

US-08-728-323A-1

Query Match 2.8%; Score 144; DB 2; Length 3489;  
 Best Local Similarity 45.1%; Pred. No. 2.5e-27;

Matches 534; Conservative 0; Mismatches 650; Indels 0; Gaps 0;

QY 1233 GGATCAGAGGCTGCTTGGAGAGCCGCTGTCAAGATGACAGCAGCAGAGAAACT 1292

Db 1578 GCAGAGACCCACAGCAGAGAGCCACAGAGCCACAGAGAGCAGAGAGCAGAGCA 1637

QY 1293 GCTGTACATTTGAATATAAGAAAGCGGAGAACTTGCAGCAGAGCAGTGTGAGCTGGA 1352

Db 1638 GCAGAGCCACAGCAGAGGAGCCACAGCAGCGGAGCCAGAGAGCGGAGCCACAGCA 1697

QY 1353 GAAGCGCCGCAAGCGCTTGTGAGCAGCAGCGCAAGAGCAGAGAGCGTTGGCTCAGCT 1412

Db 1698 GCGGAGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGAGAGCGGAGCCACAGCA 1757

QY 1413 GGAAGCGCCGAGAGAGAGAAAGAGCGGAGAGCGCCAGAGAGAGAGCGCAAGCGGCA 1472

Db 1758 GCGGAGCCACAGCAGCGGAGCCACAGCAGAGATGACAGCAGAGATGACAGCA 1817

QY 1473 GCTGAGCTGAGAAAGCAGCTGAGAAAGCAGCGGAGCTGAGCGGAGAGAGAGGA 1532

Db 1818 GCAGATGAGCAGCAGAGATGAGCAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCA 1877

QY 1533 GAGAGAGAGAGATCGAGAGCGCGGAGCCGCAAAACGGAACTGGAAGGACGCA 1592

Db 1878 GCAGATGAGCAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCA 1937

QY 1593 ACTGAATGGAAGCGGAACCGGAGACAGGAATCTCTGAATCAGAGAAACAAGAGCAGGA 1652

Db 1938 GCAGATGAGCAGCAGCAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCA 1997

QY 1653 GGGACCGTGTCTCTGAAGGCAAGAGAAAGACTTGTGAGTTGAGTTAGAGCTTGAA 1712

Db 1998 GCAGATGAGCAGCAGCAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCA 2057

QY 1713 TGACAAAAGCATCAGCTAGAGAAACCTTCAGAGATATCAGTGTGACTGCAACCCA 1772

Db 2058 GCAGATGAGCAGCAGCAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCA 2117

QY 1773 GAGGCAAGAAATTGAGAGCAAGCAAGTCTAGAGCTAAGAACTGCTGAATCAACCA 1832

Db 2118 GAGCAGCAGAGATGAGCAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCAGCA 2177

QY 1833 CTACAGCAGCAGCTTGCAGAACTCTCAGCAATCTCTGGAAGCTTATTCAGAGAAACA 1892

Db 2178 GAGTGAAGCAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCA 2237

QY 1893 GATACTCAGTACCAATTAAACAAAGTCCAGCAGAACAGTTTGTATAGACTGCTTCT 1952

Db 2238 GCAGCAGCAGCAGATGAACAGAGCAGCAGAGAGCAGCAGCAGCAGAGAGAGAGCA 2297

QY 1953 TACCTCAAAAGAGCTTGAAGCAAGAGAGCTGCGCGAGCAGCTCCGGAGCAGCT 2012

Db 2298 GCAGAGATTAGAGAGCAGAGCAGAGAGATTAGAGAGTACAGAGAGAGATTAGAGAGCA 2357

QY 2013 GACAGAGCTGAGAGAGAGAGCAGCAGCTCAAAAGCTCAGAGATTGATTTTCAACCA 2072

Db 2358 GAGAGCAGAGATTAGAGAGCAGAGCAGAGAGATTAGAGAGCAGAGAGAGATTAGAGGA 2417

QY 2073 GCTGAAGAACTGAGAGAGATATAGCAAAACAGCACTCCAGAAAGCAGAGTCCCTGGA 2132

Db 2418 GCAGAGCAGAGATTAGAGAGCAGAGCAGAGAGATTAGAGAGCAGAGAGCAGAGATTAGA 2477

QY 2133 GCGACCGGAGCTGAAGCAGAAAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCA 2192

Db 2478 GAGCAGAGAGAGAGATTAGAGAGCAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGCA 2537

QY 2193 GGAAGACCTCAGAGACGAGTTTCAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCA 2252

Db 2538 GCAAGACAGGT 2597

QY 2253 GAG 2312

Db 2598 GAG 2657

QY 2313 TGTCAAGAAAGGT 2372

Db 2658 GCAG 2717

QY 2373 GCTTTTCATCCGATCAGGT 2416

Db 2718 GGAAG 2761

## RESULT 7

US-09-298-568-1

; Sequence 1, Application US/09298568

; Patent No. 6322792

; GENERAL INFORMATION:

; APPLICANT: Kieff, Elliott D.



Query Match 2.8%; Score 144; DB 2; Length 32207;  
 Best Local Similarity 45.1%; Pred. No. 9.8e-27;  
 Matches 534; Conservative 0; Mismatches 650; Indels 0; Gaps 0;

QY 1233 GCATCAGAGGCTCTGAGAGCCCTGTCAGATGAGCAGCAGCCAGAGAGAACT 1292  
 DB 20419 GCAGAGCCACAGCAGAGAGCCAGCAGCAGCCAGCAGCAGAGCCAGCAGCA 20360  
 QY 1293 GCTGTGACATTTGAAGATGAAGCCGGAGAACTTCAGCCAGGCAATGTGAGCTTGA 1352  
 DB 20359 GCAGAGCCACAGCAGAGCCAGCAGCAGCCAGCAGCAGCCAGCAGCAGCA 20300  
 QY 1353 GAAGCCGCCAGCAGCAGCTTTGAGCAGCAGCCAGCAGCAGCAGCTGCTCAGCT 1412  
 DB 20299 GCGGAGCCACAGCAGCCAGCAGCCAGCAGCCAGCAGCCAGCAGCCAGCAGCA 20240  
 QY 1413 GAGCGCGCCAGCAGCAGAGAGAAAGCCGAGCCAGCAGCAGAGCCAGCAGCA 1472  
 DB 20239 GCGGAGCCACAGCAGCCAGCAGCCAGCAGCCAGCAGCCAGCAGCCAGCAGCA 20180  
 QY 1473 GCTGTGACATTTGAAGATGAAGCCGGAGAACTTCAGCCAGGCAATGTGAGCTTGA 1532  
 DB 20179 GCAGAGCCACAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCA 20120  
 QY 1533 GAGGAGGAAGAGATGAGCAGCAGCCAGCAGCCAGCAGCCAGCAGCCAGCAGCA 1592  
 DB 20119 GCAGAGCCACAGCAGCCAGCAGCCAGCAGCCAGCAGCCAGCAGCCAGCAGCA 20060  
 QY 1593 ACTTGAATGGGAAACGAAACCGAGACAGAACTCTGAATCAGAGAAACAGAGCAGCA 1652  
 DB 20059 GCAGAGCCACAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCA 20000  
 QY 1653 GGGCAGCCTGTCTCTGAAGCCAGAGAGAACTCTGAGATTTGAGTTGAAGCTTGA 1712  
 DB 19999 GCAGAGCCACAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCA 19940  
 QY 1713 TGACAAAACATCAGTAAAGAAATCTCAGATATCAGTGTGCACTGCAACCA 1772  
 DB 19939 GCAGAGCCACAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCA 19880  
 QY 1773 GAGGCAAGAAATGAGAGCAGAACTGTAGAGCTTAAGAAATCTGAAATCAACCA 1832  
 DB 19879 GCAGAGCCACAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCA 19820  
 QY 1833 CTTACAGCAGCAGTGTGAGAGATCTCAGCAATGCTTGAAGACTTATTCAGAGAAAC 1892  
 DB 19819 GCATGAGCAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGATGA 19760  
 QY 1893 GATACAGTGAACCACTTAAACAACTCCAGCAGAACTTGCATAGACACTCGCTTCT 1952  
 DB 19759 GCAGAGCCACAGCAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 19700  
 QY 1953 TACCTCAAAAGAGCTTGAAGCAGAAAGAGCTGGCCGAGCAGCTCCGGAGCAGCT 2012  
 DB 19699 GCAGAGCCACAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGCAGCAGCAGCA 19640  
 QY 2013 GCAGAGCCACAGCAGAGCAGCAGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCA 2072  
 DB 19639 GCAGAGCCACAGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 19580  
 QY 2073 GCTGAAGAACTGAGAGATACATACCAACAGCACTCCAGAGCAGAGGCTCCCTGGA 2132  
 DB 19579 GCAGAGCCACAGCAGAGTGAAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 19520  
 QY 2133 GGCAGCCGCACTGAAGCAGAAAGCAGAGAGAGAGCAGCAGCAGCAGCAGCAGCA 2192  
 DB 19519 GCAGAGCCACAGCAGAGTGAAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 19460  
 QY 2193 GGAAGCAGCCTGAGAGCAGATTCAGAAAGGAGCAGCAATGCTGAGCAGATGTCAGCA 2252  
 DB 19459 GCAAGAGCAGAGGTGGAAGAGCAGAAAGCAGAGCAGAGAGAGCAATTTAGAGAGGT 19400

QY 2253 GCAGAGCCACAGCAGCCCGGAAACCCACAGAGAGAGACAGTGAAGAGCAAGCAG 2212  
 DB 19399 GCAGAGCCACAGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 19340  
 QY 2313 TGTCAAGAAAGAGAGCGGAGAGAGAGCCAGCAGCAGCAGCAGCAGCAGCAGCA 2372  
 DB 19339 GCAGAGCCACAGCAGAGTGAAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 19280  
 QY 2373 GCTTTCCATCCGATCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 2416  
 DB 19279 GGAAGAGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 19236

## RESULT 9

US-08-757-669A-20/c  
 ; Sequence 20, Application US/08757669A  
 ; Patent No. 6183751  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Chang, Yuan  
 ; APPLICANT: Bohenzky, Roy A.  
 ; APPLICANT: Russo, James J.  
 ; APPLICANT: Edelman, Isidore S.  
 ; APPLICANT: Moore, Patrick S.  
 ; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS  
 ; TITLE OF INVENTION: SEQUENCES AND USES THEREOF  
 ; NUMBER OF SEQUENCES: 20  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Cooper & Dunham LLP  
 ; STREET: 1185 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: U.S.A.  
 ; ZIP: 10036  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/757,669A  
 ; FILING DATE:  
 ; CLASSIFICATION: 424  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: White, John P.  
 ; REGISTRATION NUMBER: 28,678  
 ; REFERENCE/DOCKET NUMBER: 45185-F  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 278-0400  
 ; TELEFAX: (212) 391-0525  
 ; INFORMATION FOR SEQ ID NO: 20:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 32207 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; US-08-757-669A-20

Query Match 2.8%; Score 144; DB 4; Length 32207;  
 Best Local Similarity 45.1%; Pred. No. 9.8e-27;  
 Matches 534; Conservative 0; Mismatches 650; Indels 0; Gaps 0;

QY 1233 GCATCAGAGGCTCTGAGAGCCCTGTCAGATGAGCAGCAGCCAGAGAGAACT 1292  
 DB 20419 GCAGAGCCACAGCAGCAGAGCCAGCAGCAGCCAGCAGCAGCAGCAGCAGCA 20360  
 QY 1293 GCTGTGACATTTGAAGATGAAGCCGGAGAACTTCAGCCAGGCAATGTGAGCTTGA 1352  
 DB 20359 GCAGAGCCACAGCAGAGCAGCAGCCAGCAGCCAGCAGCCAGCAGCCAGCAGCA 20300  
 QY 1353 GAAGCCGCCAGCAGCAGCTTTGAGCAGCAGCCAGCAGCAGCAGCAGCAGCAGCA 20240  
 DB 20299 GCGGAGCCACAGCAGCCAGCAGCCAGCAGCCAGCAGCCAGCAGCCAGCAGCA 20240





LOCATION: 1645..2511  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 2512..8070  
 US-08-056-200-93

Query Match 2.4%; Score 119.6; DB 1; Length 9551;  
 Best Local Similarity 44.1%; Pred. No. 1.1e-20;  
 Matches 500; Conservative 0; Mismatches 634; Indels 0; Gaps 0;

1233 GATCAGAGCTGCTTGAAGAGCCCTGTGTGATGAGATGAGCAGCAGCAGAGAAACT 1292  
 3324 GCAGAGAGAGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3383  
 1293 GCTGTGACATTTGAGATTAAGAAAGCGAGAACTTTCAGAGCAGAGCAGTGTGAGCTGGA 1352  
 3384 GCAGCTAGAGCGCGAG 3443  
 1353 GAAAGCGCGCGAG 1412  
 3444 GAGAGAGCGCGAG 3503  
 1413 GAGAGCGCGCGAG 1472  
 3504 GAGAGCGCGAG 3563  
 1473 GCTGAGAGCTGAG 1532  
 3564 GCGGAG 3623  
 1533 GAGAGAGAGAGAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1592  
 3624 GCAG 3683  
 1593 ACTTGAATGGGAAAG 1652  
 3684 GCTGAGAGCGCGAG 3743  
 1653 GGGGAGCGTGTCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1712  
 3744 CGAG 3803  
 1713 TGACAAAGAGATCAGCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1772  
 3804 GAAAGCGAG 3863  
 1773 GAGGCAAGAAATTGAG 1832  
 3864 CGAG 3923  
 1833 CTTACAG 1892  
 3924 GCAG 3983  
 1893 GATACCTCAGTACAGCAATTAAAGCAAGTCCAGAGAGAGAGAGAGAGAGAGAGAGAG 1952  
 3984 GCTGAAGCGCGAG 4043  
 1953 TACCCTTAAAG 2012  
 4044 AAGAGCGAG 4103  
 2013 GAG 2072  
 4104 CGAG 4163  
 2073 GCTGAAG 2132  
 4164 GCTGAAGCGCGAG 4223  
 2133 GGCAGCGCGAG 2192  
 4224 CGAG 4283

2193 GGAAGAGCTCAG 2252  
 4284 GCGGTGAAGCGCGAG 4343  
 2253 GAG 2312  
 4344 GCAG 4403  
 2313 TGTCAG 2366  
 4404 GCTGAAGCGCGAG 4457

# RESULT 13 US-08-800-644-93

Sequence 93, Application US/0800644  
 Patent No. 5958752  
 GENERAL INFORMATION:  
 APPLICANT: Steinert, Peter M.  
 APPLICANT: Lee, Seung-Chul  
 APPLICANT: Kim, In-Gyu  
 APPLICANT: Chung, Soo-Il  
 APPLICANT: Park, Sang-Chul  
 TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
 TITLE OF INVENTION: Methods of Using Same  
 NUMBER OF SEQUENCES: 117  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Knobbe, Martens, Olson & Bear  
 STREET: 620 Newport Center Drive, Sixteenth Floor  
 CITY: Newport Beach  
 STATE: CA  
 COUNTRY: U.S.A.  
 ZIP: 92660  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/800,644  
 FILING DATE: 14-FEB-1997  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/056,200  
 FILING DATE: 30-APR-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Fedrick, Michael F.  
 REGISTRATION NUMBER: 36,799  
 REFERENCE/DOCKET NUMBER: NIH054,001A  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (714) 760-0404  
 TELEFAX: (714) 760-9502  
 INFORMATION FOR SEQ ID NO: 93:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 9551 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1507..1644  
 FEATURE:  
 NAME/KEY: intron  
 LOCATION: 1645..2511  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 2512..8070  
 US-08-800-644-93

Query Match 2.4%; Score 119.6; DB 2; Length 9551;  
 Best Local Similarity 44.1%; Pred. No. 1.1e-20;  
 Matches 500; Conservative 0; Mismatches 634; Indels 0; Gaps 0;

1233 GATTCAGAGGCTGCTGAGAGGCGCTGCTCAGAGAGTGAAGCCAGAGAAACT 1292  
 3324 GCAGGAGGAGGCGGAGCAGCAGAGAGGCGCGAGCAGCAGAGAGAGGCGGAGCA 3383

1293 GCTGTGACATTTGAAGATTAAGAAAGCGGAGAACTTCAGCCAGGCACTGTGAGCTGGA 1352  
 3384 GCAAGCTGAGAGCGGAGCAGAGAGAGAGCGCGAGCAGCAGCTGAGCGCGAGCAGAGAGA 3443

1353 GAAGCGCGCCCAAGCGCTCTTGGAGCAGCAGCGCAAAAGACAGAGAGCGGTTGCTCAGCT 1412  
 3444 GAGAGAGCGCGAGCAGCAGCTGAGCGCGAGCAGAGAGAGAGCGCGAGCAGAGCT 3503

1413 GAGAGCGCGCGAGCAGAGAGAGAGAGAGCGGAGCGCGAGCAGAGAGAGAGAGAGAGAG 1472  
 3504 GAGGCGCGAGCAGAGAGAGAGAGAGCGCGAGCAGCAGCTGAGAGCGCGAGCAGAGCTGAG 3563

1473 GCTGAGAGCTGAGAGAGAGAGAGAGAGCGGAGCTGAGCGCGAGCAGAGAGAGAGA 1532  
 3564 GCGCGAGCAGCAGCTAGAGCTGAGCGCGAGCAGCAGCTGAGCGCGAGCAGCTGAGCGCGCA 3623

1533 GAGGAGAGAGAGAGATCAGAGAGCGCGAGCGCGCAAAACGAGAACTTGAAAGAGCAGCGACA 1592  
 3624 GAGAGAGCTGAGAGCGCGAGAGCAGCTGAGAGCGCGAGCAGCTGAGCGCGAGAGAGCA 3683

1593 ACTTGAAATGGAG 1652  
 3684 GCTGAGAGCGCGAGCAGCAGCTGAGCGCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3743

1653 GGGCAGCGGTGCTCTGAGAGCAAGAGAGAACTCTGAGATTGAGATTGAGAGCTCTGAA 1712  
 3744 CAGAGCAGAGAGAGCGCGAG 3803

1713 TGAAG 1772  
 3804 GAGAGCGCGAG 3863

1773 GAGGAG 1832  
 3864 CGAGCAG 3923

1833 CTACAGCAG 1892  
 3924 GAGAGCAG 3983

1893 GATACCTCAGTGAACAGTTAAACAGATCCAGCAGAAAGAGTTGTCATGAGAGAGAGAGAGAG 1952  
 3984 GCTGAAGCGCGAG 4043

1953 TACCTTCAAAAG 2012  
 4044 AAGACCGAG 4103

2013 GAG 2072  
 4104 CGAGCAG 4163

2073 GCTGAAG 2132  
 4164 GCTGAAGCGCGAG 4223

2133 GAGAGCGCGAG 2192  
 4224 CGAGCAGCGAG 4283

2193 GAG 2252  
 4284 GCGGCTGAAGCGCGAG 4343

2253 GAG 2312

Db 4344 GCAG 4403  
 Qy 2313 TGTGAG 2366  
 Db 4404 GCTGAAGCGCGAG 4457

RESULT 14  
 US-09-249-585A-2  
 ; Sequence 2, Application US/09249585A  
 ; Patent No. 6417002  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Horlick, Robert  
 ; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES  
 ; FILE REFERENCE: 0867/0d905  
 ; CURRENT APPLICATION NUMBER: US/09/249,585A  
 ; NUMBER OF SEQ ID NOS: 18  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 2  
 ; LENGTH: 1926  
 ; TYPE: DNA  
 ; ORGANISM: Epstein Barr Virus  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(1926)  
 ; OTHER INFORMATION: coding strand of EBNA-1 DNA  
 US-09-249-585A-2

Query Match 2.3%; Score 119.2; DB 4; Length 1926;  
 Best Local Similarity 53.4%; Pred. No. 5.2e-21;  
 Matches 250; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

Qy 1231 GTGATCAGAGGCTGCTGAGAGAGCGCTGCTCAGAGATGAGAGAGAGAGAGAGAGAGAG 1290  
 Db 515 GAG 574

Qy 1291 CTGCTGTGACATTTGAAGATTAAGAAAGCGGAGAACTTCAGAGCAGAGCAGAGCTG 1350  
 Db 575 GGGCAG 634

Qy 1351 GAGAGCGCGCCCAAGCGCTCTTGGAGCAGAGAGCGCAAAAGAGCAGAGAGAGAGAGAGAG 1410  
 Db 635 GGGCAG 694

Qy 1411 CTGAGAGCGCGCCAG 1470  
 Db 695 CAG 754

Qy 1471 CAGCTGAGCTGAG 1530  
 Db 755 GGGCAG 814

Qy 1531 GAG 1590  
 Db 815 GAG 874

Qy 1591 CACTTGAATGAG 1650  
 Db 875 CAG 934

Qy 1651 GAGGAG 1698  
 Db 935 GAGCAG 982

RESULT 15  
 US-09-050-863-2  
 ; Sequence 2, Application US/09050863  
 ; Patent No. 611411  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lao, Ying



APPLICANT: Hiang, Betty  
APPLICANT: Payan, Don  
TITLE OF INVENTION: Mammalian Protein Interaction Cloning  
TITLE OF INVENTION: System  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111-4187

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/050,863  
FILING DATE: 30-MAR-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 949-8711

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2580 base pairs

TYPE: nucleic acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: DNA

US-09-050-863-2

Query Match 2.3%; Score 119.2; DB 3; Length 2580;  
Best Local Similarity 53.4%; Pred No. 6, 2e-21;

Matches 250; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

QY 1231 GTGGATCAGAGGCTGCTTGAAGAGCCGTCGTCAGAGATGACGACGACGAGAAAGAA 1290  
DB 898 GAGAGAGGAGCAGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 957  
QY 1291 CTGCTGTGATTTGAAGTTAAGACCGGAGAACTTCAGAGGAGCACTGTGAGCTG 1350  
DB 958 GGGCAGAGGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1017  
QY 1351 GAGAAGGCGCCGCAAGGCTCTTGAAGCAGACGCAAAAGACGAGGAGCGGTTGCTCAG 1410  
DB 1018 GGGCAGAGGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1077  
QY 1411 CTGAGCGCGCCGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1470  
DB 1078 CAGGAGAGGAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1137  
QY 1471 CAGCTGAGCTGAGAAAGCAGTGAAGCAGACGAGGAGCTGAGCGGACGAGAGGAGGAG 1530  
DB 1138 GGGCAGAGGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1197  
QY 1531 GAGAGAGGAGAGATCGAGAGCGCGGAGCGCAAAACGGGAACCTGAAAAGCAGCGA 1590  
DB 1198 GAGGAGGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1257  
QY 1591 CAACCTGAATGGGAACGGAACCGGAGACAGGAACCTCCTGAATCAGAGGAACAAAGAGCAG 1650  
DB 1258 CAGGAGGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1317  
QY 1651 GAGGAGCACCTGTGCTCTGAAGGCAAGGAGGAGAACTCTGAGATTGTGAG 1698  
DB 1318 GAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1365

Search completed: March 14, 2003, 11:14:56  
Job time : 488.483 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2003, 01:40:05 ; Search time 12403 Seconds  
(without alignments)

11929.307 Million cell updates/sec

Title: US-09-674-237A-1

Sequence: 1 cggcacgagagagtgag.....gagaattcgatatcaagctt 5084

Scoring table: IDENTITY\_NUC

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: gb.ba:\*  
2: gb.hg:\*  
3: gb.in:\*  
4: gb.om:\*  
5: gb.ov:\*  
6: gb.pac:\*  
7: gb.ph:\*  
8: gb.pl:\*  
9: gb.pr:\*  
10: gb.rc:\*  
11: gb.sts:\*  
12: gb.sy:\*  
13: gb.un:\*  
14: gb.vi:\*  
15: em.ba:\*  
16: em.fun:\*  
17: em.hum:\*  
18: em.in:\*  
19: em.mu:\*  
20: em.om:\*  
21: em.or:\*  
22: em.ov:\*  
23: em.pat:\*  
24: em.ph:\*  
25: em.pl:\*  
26: em.ro:\*  
27: em.sts:\*  
28: em.un:\*  
29: em.vi:\*  
30: em.hg.hum:\*  
31: em.hg.inv:\*  
32: em.hg.other:\*  
33: em.hg.mus:\*  
34: em.hg.pln:\*  
35: em.hg.rod:\*  
36: em.hg.mam:\*  
37: em.hg.vit:\*  
38: em.sv:\*  
39: em.hgo.hum:\*  
40: em.hgo.mus:\*  
41: em.hgo.other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3721.4	73.2	3723	10	AF132478	AF132478 Mus muscu
2	3639.4	71.6	5145	10	AF132481	AF132481 Mus muscu
3	3529.4	69.4	5381	9	AF114488	AF114488 Homo sapi
4	3517.2	69.2	4025	10	AF127798	AF127798 Rattus no
5	3440.2	67.7	5287	9	AF064243	AF064243 Homo sapi
6	3103.6	61.0	3812	10	AF132672	AF132672 Rattus no
7	2975	58.5	6439	9	AF114487	AF114487 Homo sapi
8	2898.4	57.0	7247	9	AF064244	AF064244 Homo sapi
9	2047.2	40.3	4103	5	AF032118	AF032118 Xenopus l
10	1404.2	27.6	2131	9	AK027846	AK027846 Homo sapi
11	1252.6	24.6	3241	9	HSU61166	U61166 Human SH3 d
12	1110	21.8	1133	10	AF169621	AF169621 Mus muscu
13	1039.4	20.4	230097	9	AC126051	AC126053 Mus muscu
14	883	17.4	1996	9	AF180522	AF180522 Homo sapi
15	813.6	16.0	78190	9	AC106760	AC106760 Homo sapi
16	813.6	16.0	179264	2	AC023112	AC023112 Homo sapi
17	731.4	14.4	5938	9	AF033082	AF033082 Homo sapi
18	726.4	14.3	5828	9	AF248540	AF248540 Homo sapi
19	724.2	14.2	4557	9	AF182199	AF182199 Homo sapi
20	724.2	14.2	6091	9	AF182198	AF182198 Homo sapi
21	693.4	13.6	136612	2	AC115529	AC115529 Rattus no
22	692.6	13.6	180637	2	AC098552	AC098552 Rattus no
23	677	13.3	4977	10	AF132479	AF132479 Mus muscu
24	673.8	13.3	3544	6	AX428480	AX428480 Mus muscu
25	659	13.0	4447	6	AX428899	AX428899 Sequence
26	633.4	12.5	181782	2	AC116970	AC116970 Rattus no
27	556.2	10.9	11777	9	AF000312	AF000312 Homo sapi
28	556.2	10.9	100000	9	AF000193	AF000193 Homo sapi
29	556.2	10.9	114929	9	AF000050	AF000050 Homo sapi
30	556.2	10.9	151516	9	AF0000117	AF0000117 Homo sapi
31	556.2	10.9	340000	9	AF0001718	AF0001718 Homo sapi
32	553.8	10.9	2020	9	HS17581	AF064245 Homo sapi
33	510	10.0	139113	2	AC129075	AC129075 Felis cat
34	509.4	10.0	235426	2	AC121074	AC121074 Canis fam
35	508.6	10.0	178801	2	AC121075	AC121075 Canis fam
36	484.8	9.5	2017	6	AX366305	AX366305 Sequence
37	387.4	7.6	1329	9	AK021545	AK021545 Homo sapi
38	375.2	7.4	529	9	HUMY08B09	AF085873 Homo sapi
39	333.2	6.6	2873	9	AK175271	AK175271 Sequence
40	331.4	6.5	1391	9	AF001630	AF001630 Homo sapi
41	331.4	6.5	2004	9	AF038189	AF038189 Homo sapi
42	327.8	6.4	171603	9	AC012629	AC012629 Homo sapi
43	258.8	5.1	4053	9	HSU61167	U61167 Human SH3 d
44	258	5.1	747	6	AR175268	AR175268 Sequence
45	230.2	4.5	175137	2	AC115339	AC115339 Rattus no

## ALIGNMENTS

RESULT 1			
AF132478			
LOCUS	AF132478	3723 bp	RNA linear ROD 09-MAR-1999
DEFINITION	Mus musculus Esel protein mRNA, complete cds.		
ACCESSION	AF132478		
VERSION	AF132478.1	GI:4378884	
KEYWORDS	.		
SOURCE	Mus musculus.		
ORGANISM	Mus musculus.		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Scurionath; Muridae; Murinae; Mus.		
	1 (bases 1 to 3723)		
REFERENCE	Seegar,A.S., Wang,W., Bishay,J., Cohen,S. and Egan,S.E.		
AUTHORS	The EH and SH3 domain Ese proteins regulate endocytosis by linking		
TITLE	to dynamin and Eps15		

JOURNAL EMBO J. 18 (5), 1159-1171 (1999)  
 MEDLINE 99164083  
 PUBMED 10064583  
 REFERENCE 2 (bases 1 to 3723)  
 AUTHORS Sengar, A.S., Mang, W., Cohen, S., Bishay, J. and Egan, S.E.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-MAR-1999) Programs in Cancer & Blood  
 Research/Developmental Biology, The Hospital for Sick Children, 555  
 University Avenue, Toronto, ON M5G-1X8, Canada  
 Location/Qualifiers

## FEATURES

source  
 1. .3723  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 1. .3642  
 /codon\_start=1  
 /product="Ese1 protein"  
 /protein\_id="A019746.1"  
 /db\_xref="GI:4378885"

## CDS

/translation="MAQPTPTFGSLDVAIVTEBAKHDPQLSKPIAGITDDA  
 RNFPSGLPQVLAQIWAADNDGRNDQVFSIMKLIKLQGVQLPSTLPVM  
 KQDPVASSAPAGIGIASMPETAAVPMGSLPVMSPLVASPPAAEWAVP  
 GAPVIAQPLPAFAHPATWPKSSFSRSGPQLTKQKQSPVSAEPAEAVP  
 OSSRLKYQLFNHDKTMSGLTGPARTILMOSSLPOKQASIMNLSIDDDGLTA  
 EEFILAMHLIDVAMSGPLPVLPEYIPSPRRVRSQSGMSVSSSVDDLPERS  
 SEDQOEKKLPVTFEDKRENPERSVLEFRQALLQCKEQRERLAQLERAEQR  
 KEERERQAKQLELEKQLEKQLEERERERERERERERERERERERERER  
 NRQELNQRKEQBGTVLKARKTLEFELALNDKHLQGLQDIRCLATROE  
 IESTNSRELRILAEITHLQLOQLOESQOMGLRIPEKQILSDQLKQONSLDIL  
 LKRALAEKARQLEQLEDEVERETRSKQLEIDVENNOLKREIHQKQLOKQRL  
 EARLKQKQERQKLELEKQEDACRQVORQKQLEHVOQEOBRPRKPHEDLKR  
 EDSVRKEAEERAKPEMODKSRLLPHEQEPKALTOAWSTTEKGLTISAQSVK  
 VYRALYPERESHDEITIQDIDVWDESGEGLMGLKGTGMPNPAKIP  
 ENENVPKAPVTDIUSAPAKIALRETPALPVTSSBETLENNADSSSTPSSNE  
 KEBIDWDWMAQPSLTVPSAQQLQKRSAPATVAGSPVLSQGLKQVGLQAL  
 YBRAKNDNHLNFKNSDIVLEQODMMFQVQKQKFPYSYKLISGPRKSTSI  
 DTGTPESPASLKRVASPAKPAIPGEFETAMTYTESSEGDLTFQCGIIVTKKDG  
 WMTGTGKSGVPSNYRLKDESGGTGKSGTSGKKEIHOVIAVAATPEQDLT  
 APGOLILRRKNDGWEGLQARGKRIQGFNPANYVLLSPGSKITPHELPCTAV  
 OPACQVIGIMYDITANONDELAFSKQIILINLKEDPDMKKEVSGVGLFPNSYKL  
 TTDMPDSQ"

BASE COUNT 1065 a 940 c 1026 g 692 t  
 ORIGIN

Query Match 73.2%; Score 3721.4; DB 10; Length 3723;  
 Best Local Similarity 100.0%; Pred. No. 0;

Matches 3722; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 259 ATGGCTCAGTTTCCACACCTTTGGTGTAGCTTGAGTCTGGCCCAATCACTGTGAG 318  
 DB 1 ATGGCTCAGTTTCCACACCTTTGGTGTAGCTTGAGTCTGGCCCAATCACTGTGAG 60  
 QY 319 GAAAGGCCAAGCATGACGAGCTTCTAGCTGAAGCCGATACGGGATTTATTA 378  
 DB 61 GAAAGGCCAAGCATGACGAGCTTCTAGCTGAAGCCGATACGGGATTTATTA 120  
 QY 379 GGTATCAAGCGAGAACTTTTTCATCTGGTTACTCAGCGCTGCTTAGACAA 438  
 DB 121 GGTATCAAGCGAGAACTTTTTCATCTGGTTACTCAGCGCTGCTTAGACAA 180  
 QY 439 ATATGGCGCTAGCGGACATGAATTAACGATGAGAGATGATCAAGTGAATTTTCCATA 498  
 DB 181 ATATGGCGCTAGCGGACATGAATTAACGATGAGAGATGATCAAGTGAATTTTCCATA 240  
 QY 499 GCCATGAGCTTATCAAACTGAAGCTAACAAGATCAGTCCCTCCCACTGCCCT 558  
 DB 241 GCCATGAGCTTATCAAACTGAAGCTAACAAGATCAGTCCCTCCCACTGCCCT 300  
 QY 559 GTTCATGAAAACAGCAACAGTGGCTATTTCCAGTGAACCAAGATTTGATAGAGGAT 618  
 DB 301 GTTCATGAAAACAGCAACAGTGGCTATTTCCAGTGAACCAAGATTTGATAGAGGAT 360  
 QY 619 GCTAGATGCGACCACTGACAGCTGTTGCTCTGTGTCATGGGCTCCATTTCAAGTTGT 678

DB 361 GCTAGCATGCCACCACTCAGCTGTGCTCTCTGTCGCAATGGGCTCCATTCAGTTGTT 420  
 QY 679 GGAATGTCCTCACCCCTTAGTATCTTCTGCTCCCTCAGACAGAGAGCTCCCTGGCTAAC 738  
 DB 421 GGAATGTCCTCACCCCTTAGTATCTTCTGCTCCCTCAGACAGAGAGCTCCCTGGCTAAC 480  
 QY 739 GGGGCTCTCCCTGATACACGCTCTGCTGCTGGTGGGATCTCTGACAGCCACATGGCCA 798  
 DB 481 GGGGCTCTCCCTGATACACGCTCTGCTGCTGGTGGGATCTCTGACAGCCACATGGCCA 540  
 QY 799 AAGAGTCTCTCTCAGAGATCTGTGTCAGAGGTCACAACTTAACACTTAAGTACAAAG 858  
 DB 541 AAGAGTCTCTCTCAGAGATCTGTGTCAGAGGTCACAACTTAACACTTAAGTACAAAG 600  
 QY 859 GCACATCATTCAGTGTGCGCAGGCGCCCTCAGACAGACAAATGGGCTGTGCTCACTCA 918  
 DB 601 GCACATCATTCAGTGTGCGCAGGCGCCCTCAGACAGACAAATGGGCTGTGCTCACTCA 660  
 QY 919 TCAAGGCTGAAATACAGGCACTTATTCACAGCCACACAAACTATAGTGCACCTTA 978  
 DB 661 TCAAGGCTGAAATACAGGCACTTATTCACAGCCACACAAACTATAGTGCACCTTA 720  
 QY 979 ACAGGTCCTCCAGGAAGAACTTCTCATGCAATCAAGTTTACCCAGGCTCAGCTGGCT 1038  
 DB 721 ACAGGTCCTCCAGGAAGAACTTCTCATGCAATCAAGTTTACCCAGGCTCAGCTGGCT 780  
 QY 1039 TCAATATGAAATCTTTTGTGATGATCAATGAAATGAAATCTCACTGCAAGAAATTTATC 1098  
 DB 781 TCAATATGAAATCTTTTGTGATGATCAATGAAATGAAATCTCACTGCAAGAAATTTATC 840  
 QY 1099 CTAGTATGACCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1158  
 DB 841 CTAGTATGACCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900  
 QY 1159 CCAGATATCATCT 1218  
 DB 901 CCAGATATCATCT 960  
 QY 1219 AGCTCTTCTTCTGTGTGATCAGAGGCTCTCTGAGAGCGCTGTGAGAGTGAAGCAGCAG 1278  
 DB 961 AGCTCTTCTTCTGTGTGATCAGAGGCTCTCTGAGAGCGCTGTGAGAGTGAAGCAGCAG 1020  
 QY 1279 CCAGAGAAAGAACTGCTGCTGATCAATTTGAAATTAAGAAAGGGGAGAACTTGCAGCGAGC 1338  
 DB 1021 CCAGAGAAAGAACTGCTGCTGATCAATTTGAAATTAAGAAAGGGGAGAACTTGCAGCGAGC 1080  
 QY 1339 AGTGTGAGCTGAGAAAGCGCGCCGCAAGCGCTCTTGGAGCAGCAGCGCAAGAGCAGAG 1398  
 DB 1081 AGTGTGAGCTGAGAAAGCGCGCCGCAAGCGCTCTTGGAGCAGCAGCGCAAGAGCAGAG 1140  
 QY 1399 CGGTTGGCTCAGCTGAGCGCGCGGACAGAGAGAAAGAGCGGAGCGCCGAGGAGCAG 1458  
 DB 1141 CGGTTGGCTCAGCTGAGCGCGCGGACAGAGAGAAAGAGCGGAGCGCCGAGGAGCAG 1200  
 QY 1459 GAGGCAAGCGGCGAGCTGAGAGCTGAGAAACAGCTGAGAAAGCGGAGCTGAGAGCGG 1518  
 DB 1201 GAGGCAAGCGGCGAGCTGAGAGCTGAGAAACAGCTGAGAAAGCGGAGCTGAGAGCGG 1260  
 QY 1519 CAGGAGAGAGAGAGAGAGAGAGAGATGAGAGGCTGAGAGCGCGCAAAACGGGAACTG 1578  
 DB 1261 CAGGAGAGAGAGAGAGAGAGAGAGATGAGAGGCTGAGAGCGCGCAAAACGGGAACTG 1320  
 QY 1579 GAAAGGAG 1638  
 DB 1321 GAAAGGAG 1380  
 QY 1639 AACAG 1698  
 DB 1381 AACAG 1440  
 QY 1699 TTAGAAGCTCTGATATGACAAAGAAAGCATCAGCTAGAGAGAAACTTTCAGAGATACAGGTG 1758  
 DB 1441 TTAGAAGCTCTGATATGACAAAGAAAGCATCAGCTAGAGAGAAACTTTCAGAGATACAGGTG 1500

QY 1759 CGACTGGCAACCCAGAGGCAAGAAATTGAGAGCAGCAAGCTTAGAGGCTTAAGAAATT 1818  
DB 1501 CGACTGGCAACCCAGAGGCAAGAAATTGAGAGCAGCAAGCTTAGAGGCTTAAGAAATT 1560  
QY 1819 GCTGAATAATCAACCTTACAGCAGAGCTTGCAGGAATCTCAGCAAAATGCTTGAAGACTT 1878  
DB 1561 GCTGAATAATCAACCTTACAGCAGAGCTTGCAGGAATCTCAGCAAAATGCTTGAAGACTT 1620  
QY 1879 ATTCCAGAGAAACAGATACCTCAGTACAGCTTAAACAAGTCCAGCAGAAAGCTTTGAT 1938  
DB 1621 ATTCCAGAGAAACAGATACCTCAGTACAGCTTAAACAAGTCCAGCAGAAAGCTTTGAT 1680  
QY 1939 AGAGACTGCTTCTTACCTTCAAAAAGCTTGAAGCAAAAGAGCTGGCCGCGCAGCAG 1998  
DB 1681 AGAGACTGCTTCTTACCTTCAAAAAGCTTGAAGCAAAAGAGCTGGCCGCGCAGCAG 1740  
QY 1999 CTCGAGAGAGCTGAGCAGAGTGAAGAGAGACCAAGCTCAAGAGATTTGAT 2058  
DB 1741 CTCGAGAGAGCTGAGCAGAGTGAAGAGAGACCAAGCTCAAGAGATTTGAT 1800  
QY 2059 GTTTTCAACAACCAAGCTGAAGAACTGAGAGATACATGCAAAAGCAACTCCAGAAAG 2118  
DB 1801 GTTTTCAACAACCAAGCTGAAGAACTGAGAGATACATGCAAAAGCAACTCCAGAAAG 1860  
QY 2119 CAGAGGCTCCCTGAGGCGAGCGCACTGAAGCAGAAAGAGAGAGAGAGCTTGAG 2178  
DB 1861 CAGAGGCTCCCTGAGGCGAGCGCACTGAAGCAGAAAGAGAGAGAGAGCTTGAG 1920  
QY 2179 TTAGAGAGCAAAAGAGAGAGCGCTCAGAGAGAGCTTCAAGAAAGGAGCAAGCAATGCTG 2238  
DB 1921 TTAGAGAGCAAAAGAGAGAGCGCTCAGAGAGAGCTTCAAGAAAGGAGCAAGCAATGCTG 1980  
QY 2239 GAGCATGTGACAGAGAGAGAGCAGCCAGCCCGGAGAAACCCAGCAGAGAGCAGACTG 2298  
DB 1981 GAGCATGTGACAGAGAGAGAGCAGCCAGCCCGGAGAAACCCAGCAGAGAGAGCAGACTG 2040  
QY 2299 AAGAGGAGAAACAGCTGTCAGAGAAAGAGAGCGGAGAGAGAGAGAGCGGAAATGCA 2358  
DB 2041 AAGAGGAGAAACAGCTGTCAGAGAAAGAGAGCGGAGAGAGAGAGAGCGGAAATGCA 2100  
QY 2359 GACAAGCAGAGTGGCTTTTCCATCGCATAGAGAGCAGCTAAAGCTGCGCCAGCGCA 2418  
DB 2101 GACAAGCAGAGTGGCTTTTCCATCGCATAGAGAGCAGCTAAAGCTGCGCCAGCGCA 2160  
QY 2419 CCCCTGCTTACACAGAGAAAGCGCCGCTTACCACTTCTGACAGAGAGGTGTAAGAGTG 2478  
DB 2161 CCCCTGCTTACACAGAGAAAGCGCCGCTTACCACTTCTGACAGAGAGGTGTAAGAGTG 2220  
QY 2479 GTATATTACGAGGCTGTACCCCTTTGAATCCAGAACTCAGATGAGATCAGCATCCAG 2538  
DB 2221 GTATATTACGAGGCTGTACCCCTTTGAATCCAGAACTCAGATGAGATCAGCATCCAG 2280  
QY 2539 CCAGAGATATATAGTCATGTGTGATGAAGCCAGACTGGAAGCCAGAGTGGCTTGAGGA 2598  
DB 2281 CCAGAGATATATAGTCATGTGTGATGAAGCCAGACTGGAAGCCAGAGTGGCTTGAGGA 2340  
QY 2599 GAGCTGAAGGAGAGAGAGAGTGGTCCCTGCAAACTATGCAAGAAAGATTTCCAGAAAT 2658  
DB 2341 GAGCTGAAGGAGAGAGAGAGTGGTCCCTGCAAACTATGCAAGAAAGATTTCCAGAAAT 2400  
QY 2659 GAGGTTCCCACTCCAGCCAAACCAAGTGAACGATCTGACATCTGCGCCCTGCGCCAAACTG 2718  
DB 2401 GAGGTTCCCACTCCAGCCAAACCAAGTGAACGATCTGACATCTGCGCCCTGCGCCAAACTG 2460  
QY 2719 GCTCTGCTGAGAGCCCTGCTCTTTGCAAGTGAACCTTCTGAGCCCTTCAACACCC 2778  
DB 2461 GCTCTGCTGAGAGCCCTGCTCTTTGCAAGTGAACCTTCTGAGCCCTTCAACACCC 2520  
QY 2779 AACCACTGGGAGAGCTTGAAGTTCACGTCGCGCAGCAGCTCAAAAGAGAGAGAGAGAG 2838  
DB 2521 AACCACTGGGAGAGCTTGAAGTTCACGTCGCGCAGCAGCTCAAAAGAGAGAGAGAGAG 2580  
QY 2839 GACAACCTGGATACGTCGGGCGCTCAGCCCTTCTGACCCGCTACTAGTGTGCGCAGTTA 2898  
DB 2581 GACAACCTGGATACGTCGGGCGGCTCAGCCCTTCTGACCCGCTACTAGTGTGCGCAGTTA 2640  
QY 2899 CGGAGAGATCAGCTTTTACCCGAGCAGAGCAGCTGGCTCTCCCATCTCCGCTCTG 2958  
DB 2641 CGGAGAGATCAGCTTTTACCCGAGCAGAGCAGCTGGCTCTCCCATCTCCGCTCTG 2700  
QY 2959 GGCAGAGGTGAAGAGTGAAGGCTTCAAGCCGCAAGCCCTGTATCCCTGAGAGCCAAA 3018  
DB 2701 GGCAGAGGTGAAGAGTGAAGGCTTCAAGCCGCAAGCCCTGTATCCCTGAGAGCCAAA 2760  
QY 3019 AAAGCAACCACTTAAATTTTAAACAAAGTGAAGTCAACCCGCTTCTGAGAGAGAGAG 3078  
DB 2761 AAAGCAACCACTTAAATTTTAAACAAAGTGAAGTCAACCCGCTTCTGAGAGAGAGAG 2820  
QY 3079 ATGTGCTGTGTTGAGAGAGTCAAGAGTCAAGAGGCTTCCCAAGTCTTACGTAAG 3138  
DB 2821 ATGTGCTGTGTTGAGAGAGTCAAGAGTCAAGAGGCTTCCCAAGTCTTACGTAAG 2880  
QY 3139 CTCATTTTCAGGCGCCGTAAGGAAATCCAAAGCATGATCTGAGCTTGAAGAGTCT 3198  
DB 2881 CTCATTTTCAGGCGCCGTAAGGAAATCCAAAGCATGATCTGAGCTTGAAGAGTCT 2940  
QY 3199 GCTAGCTTAAAGAGAGTGGCTTCCCGGCGCCCAAGCAGCATTTCCCGAGAGAGTTT 3258  
DB 2941 GCTAGCTTAAAGAGAGTGGCTTCCCGGCGCCCAAGCAGCATTTCCCGAGAGAGTTT 3000  
QY 3259 ATTGCAATGATACATACAGAGATTTCTGAGCAGAGAGATTTTAACTTTCAAGCAGAGGAT 3318  
DB 3001 ATTGCAATGATACATACAGAGATTTCTGAGCAGAGAGATTTTAACTTTCAAGCAGAGGAT 3060  
QY 3319 GTGATTGTGTTTACCAAGAAAGATGTGATCTGTGAACGAGAGAGTGGGAGCAAGTCC 3378  
DB 3061 GTGATTGTGTTTACCAAGAAAGATGTGATCTGTGAACGAGAGAGTGGGAGCAAGTCC 3120  
QY 3379 GAGAGTCTCCCTTCTTAACTATGTGAGCTTAAAGATTTCAAGAGGCTCTGGAATCTGCGG 3438  
DB 3121 GAGAGTCTCCCTTCTTAACTATGTGAGCTTAAAGATTTCAAGAGGCTCTGGAATCTGCGG 3180  
QY 3439 AAAACAGGAGGTTTGAAGAAAGAACTGAAATTTGCGCAGTATTTCTTACGCTGCT 3498  
DB 3181 AAAACAGGAGGTTTGAAGAAAGAACTGAAATTTGCGCAGTATTTCTTACGCTGCT 3240  
QY 3499 ACTGTGCTCGAACAACCTCAGCCCTGCTCTGAGCAGCTGATTTGATTCGGAAGAAAGAC 3558  
DB 3241 ACTGTGCTCGAACAACCTCAGCCCTGCTCTGAGCAGCTGATTTGATTCGGAAGAAAGAC 3300  
QY 3559 CCAAGTGTGATGTGGAGAGAGAGACTGCAAGCTCGAGGAGAAAGAGCCAGATAGGCTG 3618  
DB 3301 CCAAGTGTGATGTGGAGAGAGAGACTGCAAGCTCGAGGAGAAAGAGCCAGATAGGCTG 3360  
QY 3619 TTTTCAAGCAATTTATGTCAAACTTCAAGCCCGGAAACAGCAAAATCAACCCCACTGAG 3678  
DB 3361 TTTTCAAGCAATTTATGTCAAACTTCAAGCCCGGAAACAGCAAAATCAACCCCACTGAG 3420  
QY 3679 CTACCAAGACCGCAGTGCAGCCAGCAGTGTGCTCAGGTATCGGAGTACAGATTACAC 3738  
DB 3421 CTACCAAGACCGCAGTGCAGCCAGCAGTGTGCTCAGGTATCGGAGTACAGATTACAC 3480  
QY 3739 GCCCAGACGATGAGAGAACTTACGCTTCAAGCAAGCCAGATATCAAGTCTTCAACAG 3798  
DB 3481 GCCCAGACGATGAGAGAACTTACGCTTCAAGCAAGCCAGATATCAAGTCTTCAACAG 3540  
QY 3799 GAGGACCGGAGCTGGTGAAGAGAGAGTCAAGTGGGCAAGTTGGGCTCTTCCCATCAAT 3858  
DB 3541 GAGGACCGGAGCTGGTGAAGAGAGAGTCAAGTGGGCAAGTTGGGCTCTTCCCATCAAT 3600  
QY 3859 TATGTAAAGCTGACCAAGACATGAGCCCAAGCAGCAATGATATATGTTGCTCATCC 3918  
DB 3601 TATGTAAAGCTGACCAAGACATGAGCCCAAGCAGCAATGATATATGTTGCTCATCC 3660  
QY 3919 CCCCTCAGGCTTTGAAGAGTCTCAAGAGAGACCACTATCTCCATATACCTGCCAAGAGGA 3978

Db 3661 CCCCTCAGGCTTGAAGTCTCTCAAGAGACCCATCTCCATATCTACCTCCAGAGGGA 3720  
 QY 3979 TGA 3981  
 Db 3721 TGA 3723

RESULT 2  
 AF132481 5145 bp mRNA linear ROD 09-MAR-1999  
 LOCUS Mus musculus Eseil protein mRNA, complete cds.  
 DEFINITION AF132481  
 ACCESSION AF132481  
 VERSION AF132481.1 GI:4378890  
 KEYWORDS  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
 AUTHORS Sengar, A.S., Wang, W., Bishay, J., Cohen, S. and Egan, S.E.  
 TITLE The EH and SH3 domain Ese proteins regulate endocytosis by linking to dynamin and Eps15  
 JOURNAL EMBO J. 18 (5), 1159-1171 (1999)  
 MEDLINE 99164083  
 PUBMED 10064583  
 REFERENCE 2 (bases 1 to 5145)  
 AUTHORS Sengar, A.S., Wang, W., Cohen, S., Bishay, J. and Egan, S.E.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-MAR-1999) Programs in Cancer & Blood Research/Developmental Biology, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, ON M5G-1X8, Canada

## FEATURES

Source  
 CDS  
 1..5145  
 /codon\_start=1  
 /product="Eseil protein"  
 /protein\_id="MAD19749.1"  
 /db\_xref="GI:4378891"

/translation="MAQPTPPGGSLDVAITVERAHNDQFLSKPIAGITGDQA  
 RNFPGSGLPVLAQIMALADNNDMDQVEFISIMKLKIKIGVOLSTTPVW  
 KOQPAVSSAPAFGIGIASMPITAVAVPMGSIPIVMSPLTSSVPAVPLAN  
 GAPVIOPLAPAPAPATWPKSSFSFSPGSLNTKOKASFPVAPAAEMAVP  
 OSSRLKYQLFNSHDKTMSGHLTGPOARTILMOSLIPOLAISIMNSLIDDOGLTA  
 BEFTILAMHLIDVAMSGOPLPYLPPEYVTPSPRRSRSSGMSVSSSVDDRLPERS  
 SEDBOPEKKLPVTTFEDKRENFERSGVELEKRAQLLEQKEDELEKRAQLEMER  
 KEREROEAKROLEKOLEKORLEKREBERKELEERRAARLEKRELEKROLEMER  
 NRRQELNORKEOGTVVLKARKTLEBELFALNDKQGLKODIRCLAPORLE  
 IESTNKSRLRIAEITTHLQOQLESQOMGLIPKQILSDQLKQONSILHPSLIT  
 LKRALEKELARQOLREOLDEVERETRSKLOEDIVYNOLKEIRHISKOLOKORSI  
 EAARLKOKEOERKSLLEKOKEDAKROARROERDKMLLHVOCBOEPKPRPHEBIRKX  
 EDSVKKEERERAKPEMODKORSLFPHQDEPAKLTQAPMTTEGPIITISQSSVKY  
 VYRRALYPPRESHDEITLQPGDIVWDSOIGBEGMIGGELGTFPANNVAKLP  
 ENEVTPPAKPVTDLTSAAPKALRETPAPLTVPITSEBSPITPNMADSSWPSSNE  
 KPEBNTMDTMAQPSLTVPASQOLRORSATFPATGSPSVLYKLQSGVRSSTSI  
 YPMRAKNDHLNFKNSDVTIVLEQDDMMFGVCOQKMPSPSYKLISGVRKSTSI  
 DTGPSPASLTKRVASPAKPAIPGEELFAMTYSSBOGDLTFQGGIVIVTKKDD  
 WMTGVGDSGVFSPSNVYRLKDSGSGTAGKSGISGKDELTAGVITASAAAGPQDLT  
 AGQGLILIRKRPGSMWESGELDARKKKQIKMFPANNVYKLISGTSKITPTELPRTAV  
 QPAVCQVIGMTDVTQNDDELAFSGQILINVLNKDPPMKGVSQVGLFSPSYVKL  
 TTDMDPSQOMGSDHLMDLPTERRKQGYIHLLVTEENYVNDLQVTELFQPLTE  
 SELLTEKEVAMI FVMWKEILMNCNKLILALRYKRMGSEKMPVKMIGDILSAQPHMQ  
 PYIRFSCOLNGAALIQOKTDEADPDFEFVKRLAMPDRCKGSLSSFLTKMOBVTYR  
 PLIKNILENTPENHPDSSHLLKHALEKAEELCSQVNEGVREKENSDELITQAVOICE  
 GLSQOLVNSVTNCLGPRKFLHSGKLYAKSKKELYGLPNDPLILITTPYSSGSG  
 DVFSPKSNLYQTKYTPFLNEVLVKLPTDPSGDEPLFHLSHIDRVYTLAESINER  
 TAVQKIKAAASELYIETEKREKAYLVARSQATIGILNANVVEGIELKCRSHGS  
 NPYCEVTMGSCSHITKTIQDTLNPKNNSQGFIFIDLEQEVLCITVEFERDQFSDDEL  
 GTEIRVADIKKQDSKGPVTCLLHHEVPTGEIIVVRDLQLPDEP"

Query Match 71.6%; Score 3639.4; DB 10; Length 5145;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 3640; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 239 ATGGCTCAGTTTCCCAACCTTTGGTGTGATGCTGTGAGCTGAGCTGAGAG 318  
 Db 1 ATGGCTCAGTTTCCCAACCTTTGGTGTGATGCTGTGAGCTGAGAGCTGAGAG 60  
 QY 319 GAAAGGCGCAAGCATGACACAGCTCTTACCTGAAAGCGGATGAGGGATTTATTA 378  
 Db 61 GAAAGGCGCAAGCATGACACAGCTCTTACCTGAAAGCGGATGAGGGATTTATTA 120  
 QY 379 GGTGATCAAGCGAGAACTTTTTCCTCAATCTGGTAACTCCAGCTGCTTACGAA 438  
 Db 121 GGTGATCAAGCGAGAACTTTTTCCTCAATCTGGTAACTCCAGCTGCTTACGAA 180  
 QY 439 ATATGGGGCTTACGAGGATGATTAATTAAGATGGAAGATGATCAAGTGAATTTTCATA 498  
 Db 181 ATATGGGGCTTACGAGGATGATTAATTAAGATGGAAGATGATCAAGTGAATTTTCATA 240  
 QY 499 GCGATGAAGCTTATCAAACTGAACTCAAGATATCAGCTCCCTCCACACTTCCCT 558  
 Db 241 GCGATGAAGCTTATCAAACTGAACTCAAGATATCAGCTCCCTCCACACTTCCCT 300  
 QY 559 GTCATGAAGACGACACCACTGAGCTATTTCCAGTGCACACAGATTTGATAGAGGAT 618  
 Db 301 GTCATGAAGACGACACCACTGAGCTATTTCCAGTGCACACAGATTTGATAGAGGAT 360  
 QY 619 GCTGACATGACACCACTGACAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 678  
 Db 361 GCTGACATGACACCACTGACAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420  
 QY 679 GGAATGCTCCACCTTATGATGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 738  
 Db 421 GGAATGCTCCACCTTATGATGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 480  
 QY 739 GGGGCTCTCCCGTCATACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 798  
 Db 481 GGGGCTCTCCCGTCATACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540  
 QY 799 AAGATGCTCTTCCACAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 858  
 Db 541 AAGATGCTCTTCCACAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600  
 QY 859 GCAATATCATTCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 918  
 Db 601 GCAATATCATTCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660  
 QY 919 TCAAGGCTGAAATACAGGAGCTTATTAACAAGCCAGCAAACTATGATGAGCACTTA 978  
 Db 661 TCAAGGCTGAAATACAGGAGCTTATTAACAAGCCAGCAAACTATGATGAGCACTTA 720  
 QY 979 ACAGGCTCCAGGCAAGCACTATTCATGCAATCAAGTTTACCCAGGCTGAGTGGCT 1038  
 Db 721 ACAGGCTCCAGGCAAGCACTATTCATGCAATCAAGTTTACCCAGGCTGAGTGGCT 780  
 QY 1039 TCAATATGAGATCTTTTGTGACATTTGATCAAGATGGAATCTCCTGAGAAATTTATC 1098  
 Db 781 TCAATATGAGATCTTTTGTGACATTTGATCAAGATGGAATCTCCTGAGAAATTTATC 840  
 QY 1099 CTAGCTATGACCTAATTTGATGTTGCTGATGCTGATGCTGCTGCTGCTGCTGCT 1158  
 Db 841 CTAGCTATGACCTAATTTGATGTTGCTGATGCTGATGCTGCTGCTGCTGCTGCTGCT 900  
 QY 1159 CCAAGATACATCTCTCTCTCTCAAGAGTTCCTGCGGAGTGGATGTCCTGATATA 1218  
 Db 901 CCAAGATACATCTCTCTCTCTCAAGAGTTCCTGCGGAGTGGATGTCCTGATATA 960  
 QY 1219 AGCTCTCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1278  
 Db 961 AGCTCTCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020

BASE COUNT  
 ORIGIN

1448 a 1298 c 1415 g 984 t

QY	1279	CCAGAGAAGAAATCGCTCTGCAATTGGAAGATAAGAAAGCCGGAGAACTTTGACCGAGGC	1338
Db	1021	CCAGAGAAGAAATCGCTCTGTGACATTGGAAGATAAGAAAGCCGGAGAACTTTGACCGAGGC	1080
QY	1339	AGTGTGAGCTTGAGAGAAAGCCCGCCCAAGGCGCTCTTGAGCAAGCCGCAAGAGACAGAG	1398
Db	1081	AGTGTGAGCTTGAGAGAAAGCCCGCCCAAGGCGCTCTTGAGCAAGCCGCAAGAGACAGAG	1140
QY	1399	CGGTTGGCTCAGCTTGAGCGCGCCGACAGAGAGAGAAAGACCGGAGCGCCACAGACAG	1458
Db	1141	CGGTTGGCTCAGCTTGAGCGCGCCCGACAGAGAGAGAAAGACCGGAGCGCCACAGACAG	1200
QY	1459	GAGGCCAAGCGGACGTGAGAGCTGGAGAGCAAGCTGGAGAAAGCAAGCGGAGACTGGAGCGG	1518
Db	1201	GAGGCCAAGCGGACGTGAGAGCTGGAGAGCAAGCTGGAGAAAGCAAGCGGAGACTGGAGCGG	1260
QY	1519	CAGGAGAGAGAGAGAGAGAGAGAGATCTGAGAGCGCGAGGCGCGCAAAACCGGAAACTG	1578
Db	1261	CAGGAGAGAGAGAGAGAGAGAGAGATCTGAGAGCGCGAGGCGCGCAAAACCGGAAACTG	1320
QY	1579	GAAAGGCAAGCGACAATTGATGGGAAACGAAACCGAGACACAGAACTCTCTAATCTAGAG	1638
Db	1321	GAAAGGCAAGCGACAATTGATGGGAAACGAAACCGAGACACAGAACTCTCTAATCTAGAG	1380
QY	1639	AACAAAGAGAGAGAGAGGACACCGTGGTCCCTGAAGGCAAGGAGGAAACACTCTGAGAGTTT	1698
Db	1381	AACAAAGAGAGAGAGAGGACACCGTGGTCCCTGAAGGCAAGGAGGAAACACTCTGAGAGTTT	1440
QY	1699	TTAGAACTCTGAATGACAAAGAAACATCAGCTAGAGAGCAAAACTTCAGAGATTCAGGTGT	1758
Db	1441	TTAGAACTCTGAATGACAAAGAAACATCAGCTAGAGAGCAAAACTTCAGAGATTCAGGTGT	1500
QY	1759	CGACTGGCAACCCAGAGGCAAGAAATTTGAGACCGAACCAAGCTTAGAGCTTAAGATT	1818
Db	1501	CGACTGGCAACCCAGAGGCAAGAAATTTGAGACCGAACCAAGCTTAGAGCTTAAGATT	1560
QY	1819	GCTGAATTCACCCACTTACAGAGCAAGCTGGCAAGAACTCAGCAAAATCCTTGGAAAGCTT	1878
Db	1561	GCTGAATTCACCCACTTACAGAGCAAGCTGGCAAGAACTCAGCAAAATCCTTGGAAAGCTT	1620
QY	1879	ATTCACAGAGAAACAGATATCTCAGTGAACCAATTGAAACAAGTCCAGACAGACAGTTTGAT	1938
Db	1621	ATTCACAGAGAAACAGATATCTCAGTGAACCAATTGAAACAAGTCCAGACAGACAGTTTGAT	1680
QY	1939	AGAGACTCGCTTTTACCTCAAAAGAGCCTTGAAGCAAGAGAGCTGGCCCGGACAG	1998
Db	1681	AGAGACTCGCTTTTACCTCAAAAGAGCCTTGAAGCAAGAGAGAGCTGGCCCGGACAG	1740
QY	1999	CTCCGGAGACAGCTGAGACAGAGTGTGAGAGAGAACCAAGTTCAAACTCTGACAGAGATTGAT	2058
Db	1741	CTCCGGAGACAGCTGAGACAGAGTGTGAGAGAGAACCAAGTTCAAACTCTGACAGAGATTGAT	1800
QY	2059	GTTTTCACAAACAACGCTGAGAGAACTGAGAGATATCATATGCAAAACAGCAACTCCAGAA	2118
Db	1801	GTTTTCACAAACAACGCTGAGAGAACTGAGAGATATCATATGCAAAACAGCAACTCCAGAA	1860
QY	2119	CAGAGGTCCTTGAGAGAGCGCGCATTTGAACAGAAAGACAGAGAGAGAGAGGCTTGAG	2178
Db	1861	CAGAGGTCCTTGAGAGAGCGCGCATTTGAACAGAAAGACAGAGAGAGAGAGGCTTGAG	1920
QY	2179	TTTAGAGAAAGCAAAAGAAAGACGCTCAGAGACAGATTCAAGAAAGGGAACAAGATATGCTG	2238
Db	1921	TTTAGAGAAAGCAAAAGAAAGACGCTCAGAGACAGATTCAAGAAAGGGAACAAGATATGCTG	1980
QY	2239	GAGCATGTGACAGAGAGAGACAGCCAGCCCGCCGAGAAACCCACAGAGAGAGACAGACTG	2298
Db	1981	GAGCATGTGACAGAGAGAGACAGCCAGCCCGCCGAGAAACCCACAGAGAGAGACAGACTG	2040
QY	2299	AAGAGGAAAGACAGTGTCTAGAGAAAGAGAGCGGAAAGAGAGACCAAGCTGGAAATGCA	2358
Db	2041	AAGAGGAAAGACAGTGTCTAGAGAAAGAGAGCGGAAAGAGAGACCAAGCTGGAAATGCA	2100
QY	2359	GACAAAGACAGAGTCCGCTTTTTCATCCGCATCTAGAGACCAAGCTTAAGTGGCCACCCAGCA	2418

Db	2101	GACACAGAGAGTGGCTTTTCCATCCGATCGAGGCCAGGCTGACCCACGAGCA	2160
Qy	2419	CCCTGGCTTACCAAGAGAAAGCCCGCTTACCATTTCTGCACAGAGAGTGTAAAGTG	2478
Db	2161	CCCTGGCTTACCAAGAGAAAGCCCGCTTACCATTTCTGCACAGAGAGTGTAAAGTG	2220
Qy	2479	GTTATTATCCGACGGCTGATACCCCTTTGAAATCCAGAAATCGACAGATGACATCCAG	2538
Db	2221	GTTATTATCCGACGGCTGATACCCCTTTGAAATCCAGAAATCGACAGATGACATCCAG	2280
Qy	2539	CCAGAGATATTAGTCAATGGTGTATGAAAGCCAGACTGAGAGCCAGATGGCTTGAAGA	2598
Db	2281	CCAGAGATATTAGTCAATGGTGTATGAAAGCCAGACTGAGAGCCAGATGGCTTGAAGA	2340
Qy	2599	GAGCTGAAGGGAAGACGGATGTGTTCCCTGCAAACTATGCAAAAAGATTCCGAAAAT	2658
Db	2341	GAGCTGAAGGGAAGACGGATGTGTTCCCTGCAAACTATGCAAAAAGATTCCGAAAAT	2400
Qy	2659	GAGGTTCCCACTCCAGCCAAACAGGACCGATCTGACATCGCCCGCCCGCCAAATG	2718
Db	2401	GAGGTTCCCACTCCAGCCAAACAGGACCGATCTGACATCGCCCGCCCGCCAAATG	2460
Qy	2719	GCTTGGGTAGACCCCTGCTCTTTGCCAGTACCTCTTGACACCTTCACAACCCC	2778
Db	2461	GCTTGGGTAGACCCCTGCTCTTTGCCAGTACCTCTTGACACCTTCACAACCCC	2520
Qy	2779	AACAACTGGGACACTTCAGTTCCACGTGGCCGACAGCTCAACGAGAGCCAGAAACG	2838
Db	2521	AACAACTGGGACACTTCAGTTCCACGTGGCCGACAGCTCAACGAGAGCCAGAAACG	2580
Qy	2839	GACAACTGGGATACGTGGGGGTGACGCTTCCTGACCCGTACCTAGTGCTGGGACAGTTA	2898
Db	2381	GACAACTGGGATACGTGGGGGTGACGCTTCCTGACCCGTACCTAGTGCTGGGACAGTTA	2640
Qy	2899	CGGACAGATACAGCTTTTACCCGACACAGGCACTGGCTCTCCCTCCACTCCGCTCTG	2958
Db	2641	CGGACAGATACAGCTTTTACCCGACACAGGCACTGGCTCTCCCTCCACTCCGCTCTG	2700
Qy	2959	GGCCAGGATGAAAAGGTGGAAGGGCTACAAAGGCAACCTGTATCTCTGAGAGCCAAA	3018
Db	2701	GGCCAGGATGAAAAGGTGGAAGGGCTACAAAGGCAACCTGTATCTCTGAGAGCCAAA	2760
Qy	3019	AAACACACCACTTAAATTTTAAACAAGTCATACGGTCTGGAAACAGCAAGAC	3078
Db	2761	AAACACACCACTTAAATTTTAAACAAGTCATACGGTCTGGAAACAGCAAGAC	2820
Qy	3079	ATGAGTGGTGTGGAGAAGTTCAGAGTTCAGAAAGGTTGGTCCCAAGCTCTTACGTGAAA	3138
Db	2821	ATGAGTGGTGTGGAGAAGTTCAGAGTTCAGAAAGGTTGGTCCCAAGCTCTTACGTGAAA	2880
Qy	3139	CTCATTTAGGGCCCGTAGAGAAATCCACAGCATGATATGCGCCCTATCTGAAGTCTT	3198
Db	2881	CTCATTTAGGGCCCGTAGAGAAATCCACAGCATGATATGCGCCCTATCTGAAGTCTT	2940
Qy	3199	GCTAGTCTTAAAGAGAGTGGTTCCTCCCGGCCGCAACCGACCATTCGCGAAGAAAGTTT	3258
Db	2941	GCTAGTCTTAAAGAGAGTGGTTCCTCCCGGCCGCAACCGACCATTCGCGAAGAAAGTTT	3000
Qy	3259	ATTGCCATGTACATATCGAGAGTTCTGACAAAGAGATTTTAACTTTTACGAAAGGGAT	3318
Db	3001	ATTGCCATGTACATATCGAGAGTTCTGACAAAGAGATTTTAACTTTTACGAAAGGGAT	3060
Qy	3319	GTGATTGTGTACCAAGAAAGATGTGACTGTGTGACGCGAAACGCTGGGCGCAAGTCC	3378
Db	3061	GTGATTGTGTGTACCAAGAAAGATGTGACTGTGTGACGCGAAACGCTGGGCGCAAGTCC	3120
Qy	3379	GGAAGTCTTCCCTTATCTATGTGAGGCTTAAAGATTACAGAGGGCTCTGGAACTCTGGG	3438
Db	3121	GGAAGTCTTCCCTTATCTATGTGAGGCTTAAAGATTACAGAGGGCTCTGGAACTCTGGG	3180
Qy	3439	AAAAAGGAGATTGAGAAAAAAACCTGAAGTGGCCAGGTTATGCTTCTTACGCTGCT	3498



Db 3181 AAAACAGGAGTTTAGGAAAAAACTGAAATTGCCAGTTATTGCTTCTTCAAGCTGCT 3240

Qy 3499 ACTGTCCTCCGACCACTACCTCTGCTCTCTGAGCACTGATTTCTGATCCGGAAGAAAC 3558

Db 3241 ACTGTCCTCCGACCACTACCTCTGCTCTCTGAGCACTGATTTCTGATCCGGAAGAAAC 3300

Qy 3559 CCGAGGTGATGTTGGGAGGAGAACTGCAAGCTCGAGGGGAAAAAGCCAGATAGGCTG 3618

Db 3301 CCGAGGTGATGTTGGGAGGAGAACTGCAAGCTCGAGGGGAAAAAGCCAGATAGGCTG 3360

Qy 3619 TTTCAGCAATTATGTCTAACTTTAAGCCCGGAACAAGCAAAATCAACCCCACTGAG 3678

Db 3361 TTTCAGCAAAATTATGTCTAACTTTAAGCCCGGAACAAGCAAAATCAACCCCACTGAG 3420

Qy 3679 CTACCCAGACCCGAGTGCAGCCAGAGTGTGCTGAGTATGCGGATGTACGATTACACC 3738

Db 3421 CTACCCAGACCCGAGTGCAGCCAGAGTGTGCTGAGTATGCGGATGTACGATTACACC 3480

Qy 3739 GCCCAGAACGATGAGCAACTAGGCTTCAGCAAGGCGAGATCATCAAGCTCCCAAGAG 3798

Db 3481 GCCCAGAACGATGAGCAACTAGGCTTCAGCAAGGCGAGATCATCAAGCTCCCAAGAG 3540

Qy 3799 GAGGACCCGAGTGTGAGAAAGAGAGTCACTGAGGCAAGTTGGCTTTCCCATCAAT 3858

Db 3541 GAGGACCCGAGTGTGAGAAAGAGAGTCACTGAGGCAAGTTGGCTTTCCCATCAAT 3600

Qy 3859 TATGTAAGCTGACCAAGCATGAGCCCGAGCAGCAATG 3899

Db 3601 TATGTAAGCTGACCAAGCATGAGCCCGAGCAGCAATG 3641

RESULT 3

LOCUS AF114488 5381 bp mRNA linear PRI 16-JUL-2002

DEFINITION Homo sapiens interseccion short isoform (ITSN) mRNA, complete cds.

ACCESSION AF114488

VERSION AF114488.1 GI:4808824

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 5381)

AUTHORS Pucharcos, C., Fuentes, J.J., Cases, C., de la Luna, S., Alcantara, S., Arbones, M.L., Soriano, E., Estivill, X. and Pritchard, M.

TITLE Alu-splice cloning of human interseccion (ITSN), a putative multivalent binding protein expressed in proliferating and differentiating neurons and overexpressed in Down syndrome

JOURNAL Eur. J. Hum. Genet. 7 (6), 704-712 (1999)

MEDLINE 99415290

PUBMED 10482960

REFERENCE 2 (bases 1 to 5381)

AUTHORS Pucharcos, C., Fuentes, J.J., Pritchard, M. and Estivill, X.

TITLE Direct Submission

JOURNAL Submitted (16-DEC-1998) Medical and Molecular Genetics Center, Cancer Research Institute, U/Hospitalet de Llo., Avia, Castelldefels km. 2,7, Barcelona 08907, Spain

FEATURES

source

1..5381

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="21"

/map="21q22.1-q22.2"

1..5381

/gene="ITSN"

/gene="ITSN"

269..3931

/gene="ITSN"

/codon\_start=1

/product="interseccion short isoform"

/protein\_id="AAD29953.1"

/db\_xref="GI:4808825"

/translation="MAQFPPTPGSLDIWATVEBRAGHDOFSLKISGFTGDOA RNFPGSGLPVLAQIWAIDMNMNDGMDQVEFIANKIKLKIGYQPLSALPVM

Qy 7 GAGGAGAGTGGAGGCGCGGAGAGGCGCGGAGCTTGTCTGCTTCTGATGAGGCGGCT 66

Db 27 GAGGAGAGTGGAGGCGCGGAGAGGCGCGGAGCTTGTCTGCTTCTGATGAGGCGGCT 86

Qy 67 CGCAAGGAGGATCCGAGGCGGCTCCGAGACCGCGGAGCGAGCGAGCGGCGGCGG 126

Db 87 CGCAAGGAGGATCCGAGGCGGCTCCGAGACCGCGGAGCGAGCGGCGGCGGCGG 135

Qy 127 GGGATGTTGGCGGCGGCTGCGGAGTGGCGCTTCTGCG -CGGCGTGGCGGCTGCACTGA 185

Db 136 GGGATGTTGGCGGCGGCTGCGGAGTGGCGCTTCTGCG -CGGCGTGGCGGCTGCACTGA 195

Qy 186 TTTGTGAGGAGGCGGCGGCGGCGGCGGCGGAGTGGCGGCTGAGTGGCGGAGTGA 245

Db 196 TTTGTGAGGAGGCGGCGGCGGCGGCGGCGGAGTGGCGGCTGAGTGGCGGAGTGA 255

Qy 246 ACGTAAATGAACCATGCTGCTTCCACACCTTTGCTGCTGAGCTGAGTCTGGG 305

Db 256 AAGTAAATGAACCATGCTGCTTCCACACCTTTGCTGCTGAGCTGAGTCTGGG 315

Qy 306 CATACCTGTGAGGAGGAGGCGGAGGAGTGGCGGAGTGGCTTGGCTGAGGCGGAGTGC 365

Db 316 CATACCTGTGAGGAGGAGGAGGAGGAGTGGCGGAGTGGCTTGGCTGAGGCGGAGTGC 375

Qy 366 GGGATTTTACTGCTGATCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 425

Db 376 TGGATTTTACTGCTGATCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 435

Qy 426 TGTCTTGCACAAATATGCGGCTGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 485

Db 436 TGTCTTGCACAAATATGCGGCTGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 495

Qy 486 GGAATTTTCCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 545

Db 496 GGAATTTTCCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 555

Qy 546 CACACTTCCCTGCTCATGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 605

Db 556 TGCACTTCCCTGCTCATGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 615

Qy 606 TATGAGGAGGATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 665

Db 616 TATGAGGAGGATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 675

Qy 666 CATTCGAGTGTGAGGATGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 725

BASE COUNT 1629 a 1192 c 1304 g 1256 t

ORIGIN

Query Match 69.4%; Score 3529.4; DB 9; Length 5381;

Best Local Similarity 83.7%; Pred. No. 0;

Matches 4278; Conservative 0; Mismatches 766; Indels 69; Gaps 22;

KQOPVAISSAPARWGKGIASMPRLTAVAPVPMGSIPIVGMSPITLSSVPTAIVPLAN

GAPVLOPUPAPAPATLIPKSSSRSSPSGSLTKQAKQSPVAVSPAPAVAP

QSSRLKYROLFNSHDKTMSGLTGPARTIMQSSLPQQLSVISNLSIDDDGLTA

BERILAMHLIDVAMSGOPLPVLPPEYIPPSRRVRVSGSIVISVSDGLPEPV

LEDEOQOLREKTLPTVEDKRENFERNLEKROALIEQOKRELAQLEABOE

KREEREOERKROLEKORELEKRENFERNLEKROALIEQOKRELAQLEABOE

RNRROBLNORNMDEEDIVLAKAKKTLFEELALNDKHOGLQLODRGLTTORO

ELBSTKRSRELRNLAETTHLOOQLOESQMLLEKILPEKOLNOLKONOSLHRSIV

TLKRALBAKELAROLRDOLDEVEKTRSKLEIDIPNQLKELERHKOLOKOS

MEARLKOKEORERKILIELEKOEORERKQLEKOWLEHVOODEHOREKULIEEKL

KREESVKKDKEBKQOEADKLELFOHOPAPAPVAPWSTAEKGLTISAEVY

KVYVYRALLYPESRSHDEITIOGDIIVMKGWVBSOTGEBMGLGELTKGWFPA

NAVEKIENEBEVPAPVPTVSTAPAPKALBETPAPALVTSSESTPENNMADSSST

MPSTKEPTEOTMDMAAPSLTVPASQOLBORSAFPTATGSSPSVYGOQKEVE

GLOAQILPWRKNDNLNPNKQDVTIVLEQDDMMFEGVQOKMPPKSYKLISGP

ILRKSTMSDSSSPASLRVASPAKPVASBEMFTAMTYSBQGLTFPOQGVIL

VTKQDDMTWGTGVDAVFPENYVRLNDSBSGAGKGSJGKRPALQVIASTAT

GPEQLTLAPGOLILRKKNPGMWBEGLDARKKROIGWFPANVYKLSPGSKITPT

EPKSTALAAVCQVIMYDYTAQNDELAFNKGQIIVNLKEDPMMWKGVEVNGVGLF

PSNYVRLTITMDPSQO"





QY 2865 GCGCTTCTGACCCGTAAGTGTGCGCAGATTACGAGATCAGCCCTTAAACCCACG 2924  
 DB 2896 GCGCTTCTGACCCGTAAGTGTGCGCAGATTACGAGATCAGCCCTTAAACCCACG 2955  
 QY 2925 CAGAGCCATCTGCTCTCCCACTCTCCGCTCTGCGGCGAGGTTGAAAGTGAAGGCT 2984  
 DB 2956 CAGAGCCATCTGCTCTCCCACTCTCCGCTCTGCGGCGAGGTTGAAAGTGAAGGCT 3015  
 QY 2985 ACAAGCGCAAGCCCTGATCCCTGAGAGCCAAAAGCAACCTTAATTTTAAACA 3044  
 DB 3016 ACAAGCGCAAGCCCTGATCCCTGAGAGCCAAAAGCAACCTTAATTTTAAACA 3075  
 QY 3045 AAGTGAAGTCATACCGTCTGGAAGACAGACATGTGTGTTTGAAGATTTCAAG 3104  
 DB 3076 AAGTGAAGTCATACCGTCTGGAAGACAGACATGTGTGTTTGAAGATTTCAAG 3135  
 QY 3105 TCAGAGGGTGTGTTCCCAAGTCTACGTAACCTTAATTTTCAAGGCTTGAAGAAATC 3164  
 DB 3136 TCAGAGGGTGTGTTCCCAAGTCTACGTAACCTTAATTTTCAAGGCTTGAAGAAATC 3195  
 QY 3165 CACAGCATGATACCTGAGCCCTGTAAGTCTGCTAGTCTAAAGAGTGGCTTCCG 3224  
 DB 3196 TACAGCATGATACCTGAGCCCTGTAAGTCTGCTAGTCTAAAGAGTGGCTTCCG 3255  
 QY 3225 GCGCCGCAAGCCAGCCATTCGCGGAGAGATTTATTCATGTACATACGAGATTC 3284  
 DB 3256 AGCAGCCAGCCAGCCATTCGCGGAGAGATTTATTCATGTACATACGAGATTC 3315  
 QY 3285 TGAGCAAGAGATTTTAACTTTCAGCAAGGGGATGTATTTGTGTTACCAAGAAAGATG 3344  
 DB 3316 TGAGCAAGAGATTTTAACTTTCAGCAAGGGGATGTATTTGTGTTACCAAGAAAGATG 3375  
 QY 3345 TGAATGTGTGACGGAACGCTGCGGACAGATCCGAGTCTTCCCTTAACTATGTAG 3404  
 DB 3376 TGATGTGTGACGGAACGCTGCGGACAGATCCGAGTCTTCCCTTAACTATGTAG 3435  
 QY 3405 GCTTAAAGATTCAGAGGCTCTGGAACCTGCTGGAAGAAACAGGAGTTTGGAAAAAAC 3464  
 DB 3436 GCTTAAAGATTCAGAGGCTCTGGAACCTGCTGGAAGAAACAGGAGTTTGGAAAAAAC 3495  
 QY 3465 TGAATTTGCGCAGATTTATGCTTCTTACGCTGCTTACGCGCAACCTACCTGCG 3524  
 DB 3496 TGAATTTGCGCAGATTTATGCTTCTTACGCTGCTTACGCGCAACCTACCTGCG 3555  
 QY 3525 TCTGTGCGCAGCTGATTTCTGATCCGGAAGAAAGAACCCAGGTGATGTGTGGAAGAGAT 3584  
 DB 3556 CCTGTGTGCGCAGCTGATTTCTGATCCGGAAGAAAGAACCCAGGTGATGTGTGGAAGAGAT 3615  
 QY 3585 GGAAGCTGAGGGAAGAAACGCGGATAGGCTGTTTCCAGCAAAATTAATGTAACCTTCT 3644  
 DB 3616 GGAAGCTGAGGGAAGAAACGCGGATAGGCTGTTTCCAGCAAAATTAATGTAACCTTCT 3675  
 QY 3645 AAGCCCCGGAACAGCAAAATCAACCCAGTACGAGCAAGACGCGAGTCAAGCCAGC 3704  
 DB 3676 AAGCCCCGGAACAGCAAAATCAACCCAGTACGAGCAAGACGCGAGTCAAGCCAGC 3735  
 QY 3705 AGTGTGCGCAGGTTGATGGGATGATCAATTAACCGCCAGAACGATGACGAATAGCTT 3764  
 DB 3736 AGTGTGCGCAGGTTGATGGGATGATCAATTAACCGCCAGAACGATGACGAATAGCTT 3795  
 QY 3765 CAGCAAGGCGAGATCATCAAGCTCTCAACAAAGAGAGCCGCGAGTGTGGAAGAGGA 3824  
 DB 3796 CAGCAAGGCGAGATCATCAAGCTCTCAACAAAGAGAGCCGCGAGTGTGGAAGAGGA 3855  
 QY 3825 AGTCAATGGAAGATTTGAGGCTCTTCCATCCAAATTAATGTAAGTGAACAGACATGGA 3884  
 DB 3856 AGTCAATGGAAGATTTGAGGCTCTTCCATCCAAATTAATGTAAGTGAACAGACATGGA 3915  
 QY 3885 CCGCAGCCAGCAATGATATATTTGTCATCCTCCCTCAGGCTTGAAGTCTCTCAA 3944  
 DB 3916 CCGCAGCCAGCAATGATATATTTGTCATCCTCCCTCAGGCTTGAAGTCTCTCAA 3975  
 QY 3945 GAGACCCACTATCCCATATCACTGCGCCAGAGGATGATGGAGATGACGCTTATCATG 4004

DB 3976 GAGACCCACTATCCCATATCACTGCGCAGAGGATGATGGAGATGACGCTTATCATG 4035  
 QY 4005 TGACTTGCAGATGATATCACTATGCTCTTGTGATGAGAAACATCACTGAGAGGATTT 4064  
 DB 4036 TGACTTGCAGATGATATCACTATGCTCTTGTGATGAGAAACATCACTGAGAGGATTT 4095  
 QY 4065 ACCATATTGACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4124  
 DB 4096 ACCATATTGACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4155  
 QY 4125 AAGC - AATGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4182  
 DB 4156 GAGCAAAATTAACAAAACACAGAGGATGATGATGATGATGATGATGATGATGATGATGAT 4215  
 QY 4183 AGACTGACCGG - CCGGCTCTCAACAGGCGCTTCAATGTTTAAATTTTAAATTTTAAATTT 4241  
 DB 4216 AATTTGACTTTTCCCGCCTTGTGACAGGCTTCAATGTTTAAATTTTAAATTTTAAATTT 4275  
 QY 4242 TGTGATTTTAAAGCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 4301  
 DB 4276 TATATTTTAAAGCTTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 4335  
 QY 4302 TACAAAACCCACTATCAAGAGTGTGCTGCGGACGATTTAAATGCTGTTCCGG 4361  
 DB 4336 TGCAAAAAGCCACTATCAAGGATG - - - - CTGATGTCTATTTAAATTTTCCAA 4390  
 QY 4362 CGTACCGTAACCTGAGAGCTGCTGATCTTTCGCTTTCAGTGTGCTTCCACACAT 4421  
 DB 4391 ATGTCATTAATCTGAGACTGTGATGATTTTTCATTTTGTCCAGTGTACCACTAAT 4450  
 QY 4422 TGTGATTTTGGGCTGTTCC - - - - TGCGTAGACACAGAGAGATGGGTATCTGT 4477  
 DB 4451 TGTGATTTTGGGCTTTCCTTTCATTAAGATGACAGAGATGATGATGATGATGATGAT 4510  
 QY 4478 TTTTAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4537  
 DB 4511 TTTTAAAGCTTAA - GATGAGCCCAATTAAAGCAAGGTTTGTGCTTGTGTGTGT 4569  
 QY 4538 ATCAAGTGTAC - TGTGCGCATGATGATGATGATGATGATGATGATGATGATGATGAT 4595  
 DB 4570 ATCAAGTGTAC - TGTGCGCATGATGATGATGATGATGATGATGATGATGATGATGAT 4629  
 QY 4596 GCGTAAACCCAGCAGCTGTACAGTGTCTCATCTACT - GCATCAATTTTACTTTGAC 4653  
 DB 4630 TGGCAAGCTTAACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4689  
 QY 4654 AGTGAAGCTTGAAGCAGTGAAGAGACCTGTGTGTTTGTGCTGCTGATGATGATGATGAT 4712  
 DB 4690 AGTGAAGCTTGAAGCAGTGAAGAGACCTGTGTGTTTGTGCTGCTGATGATGATGATGAT 4749  
 QY 4713 AGTGTGCGGCTGTTTGTGTTTATTTTCAATCTGCAATGTTTCAACACATTAAGTAG 4772  
 DB 4750 GGTGATGTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 4809  
 QY 4773 TAAAGCCCAAGTCCAGGCGTTACGATCATCACTGATCCACCGTC - - - - TTAAGTCTC 4827  
 DB 4810 GTAGACCAAC - ACCATGAGGCTGTACGATCAAGATATCCACAGCTCTTTTATCTTC 4868  
 QY 4828 TGTAGAGTGAAG - TTTATTCAGTGTGTTTATGGA - - - - TATCTGAACAGATTAAT 4881  
 DB 4869 TGTAGAGTGAAGTGTGTTTATTCAGTGTGTTTATGGA - - - - TATCTGAACAGATTAAT 4928  
 QY 4882 CTTCTGACAAAGAAATGATATGAAAGTCTCCCTGCAATTAATTTCCAGTGTTAACT 4941  
 DB 4929 TTTCTGACAAAGAAATGATATGAAAGTCTCCCTGCAATTAATTT - CCAATGTTTAACT 4987  
 QY 4942 TTTTAACTGACTGTGCGGCTTGTACAGATTAATTAATTAATGCGCTGCTGCTGCGG 5001  
 DB 4988 TTTTAACTGACTGTG - GGAATTTTCAAGATTAATTAATTAATGAGCTCATGCTGCTGCT 5046  
 QY 5002 TGTGTTTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5058

Db 5047 TGTGTGTT-AGATATGCTGTAGCTGAAGCCCTGTTGCTTTTAAACACTAGTTGGAAGC 5105  
 QY 5059 CCGGAAGAGATT 5071  
 Db 5106 TCTCATATAAAT 5118

RESULT 4  
 AF127798  
 LOCUS AF127798  
 DEFINITION Rattus norvegicus EH- and SH3-domain containing protein EHSN1 mRNA.  
 ACCESSION AF127798  
 VERSION AF127798.1 GI:4835852  
 KEYWORDS  
 SOURCE Rattus norvegicus.  
 ORGANISM Rattus norvegicus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 4025)  
 Okamoto, M., Schoch, S. and Sudhof, T.C.  
 EHSN1/intersectin, a protein that contains EH and SH3 domains and  
 binds to dynamin and SNAP-25. A protein connection between  
 exocytosis and endocytosis?  
 J. Biol. Chem. 274 (26), 18446-18454 (1999)  
 99303609  
 10373452  
 2 (bases 1 to 4025)  
 Okamoto, M., Schoch, S. and Sudhof, T.C.  
 Direct Submission  
 Submitted (11-FEB-1999) Center for Basic Neuroscience and HHMI,  
 Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX  
 75235, USA

FEATURES  
 source  
 1..4025  
 /organism="Rattus norvegicus"  
 /db\_xref="taxon:10116"  
 41..3694  
 /product="EH- and SH3-domain containing protein EHSN1"  
 /protein\_id="AAD30271.1"  
 /translation="MAQPTTPFGSLIDIMAITVEERAKHDOQFSLKPIISGITDQA  
 /db\_xref="GI:4835853"  
 /translation="MAQPTTPFGSLIDIMAITVEERAKHDOQFSLKPIISGITDQA  
 RNFPSGSLPQVLAQIMALADMKGRDQVPSIANKLILKIQGLQPLPAPVPM  
 KOCPAISGAPAPGIGMAEMPLTAVAPVPMPSIPIVGMGSGMSVISSSADQRLPEPS  
 GAPPIYQIPAPAHAPATLPRKSSFSRSGPSQNTXQKQSPVAPASAPAAEMAVP  
 QSSRLKTYRQLENSHDKTSGHLTGPOARTILMOSLPQQLASLNNLSDIDQDKLTA  
 EFFILAMHLDIVAMSGQPLPVLPEPIYIPSRVRSGMSVISSSADQRLPEPS  
 SEDDOVEKRLPVTEDKKEKRENERGNIIEKREKQALIQKREKRELEAOLREOER  
 KEREOERKROLEKREKRENERGNIIEKREKQALIQKREKRELEAOLREOER  
 NROELTORNDQGIIVLKARRTLELEALNDKHOLGKQIDIRCLATQROER  
 IESTKSRRLTAETTHLQOQIQESQOMGLIPKQILSLDKRKYQONSIRHSLT  
 LKRALEKELARQOLREOLDEVEKTRSLQLEIDVNNOLKLEIRHSHKQQLQKRSI  
 EAERLQKQERKSELEKQKREKQREKQREKQREKQREKQREKQREKQREKQREKQ  
 EDSVKKKEAEERKSELEKQKREKQREKQREKQREKQREKQREKQREKQREKQREKQ  
 YRYALYPPESRSHDEITQPGDIIVKMGWDSQTEGEGGEPKGTGMPFANVA  
 EKIPENELPTPAKPVTDLTSAAPKALRETPAPLPVSSRSPTPNMMAPESSWPS  
 STNEKRETTNNMTMAQESLTTPSAQLRQSAFTPATRATGSPVPVQGEKEKGLQ  
 AALYTPWRKKNHNLNFKSDVITVLEODQKREKQREKQREKQREKQREKQREKQREKQ  
 STSIDGPTPEASLSLRVASPAKPAIPDEEFVAVTYESSHGDLTQOQHVIVTK  
 KQGDWMTGIVGEGTGVPSNVYAKPAIPDEEFVAVTYESSHGDLTQOQHVIVTK  
 QTLAPGQILIRKKNPGGMEGELQARKKROIEMFANPYKILSPGSKITPTLP  
 KTAIVAPCOVIGMYDYDTAQNDELAFSGQIIVLSKEDPDMMKGEVSGQGLPSPN  
 YKLTDDMPSSQ"

BASE COUNT 1147 a 1009 c 1123 g 746 t  
 ORIGIN  
 Query Match 69.2%; Score 3517.2; DB 10; Length 4025;  
 Best Local Similarity 93.4%; Pred. No. 0;  
 Matches 3734; Conservative 0; Mismatches 243; Indels 21; Gaps 5;  
 QY 219 GAGATGAGCGGTGATGAGCAAGGTGAAGTAAATAGAACCATGATGCTTCCACACACC 278

Db 1 GAGATGAGCGGTGATGAGCAAGGTGAAGTAAATAGAACCATGATGCTTCCACACACC 60  
 QY 279 TTTGCGTGTAGAGCTGTGATGTCTGGCCATACCTGTGAGAGAAAGGCCAAGCATGCA 338  
 Db 61 TTTTGTGGAGAGCTGTGACATCTGGGCATACCGTGGAGAGAAAGGCCAAGCATGCA 120  
 QY 339 GCAATTCCTTAGCCTGAAGCCGATAGCGGATTTATCTGTGTATCAGCGAGAACTT 398  
 Db 121 GCAGTTCAGAGTTTGAAGCCGATATCTGATTCATCATCGTGTATCAGCGAGAACTT 180  
 QY 399 TTTTTCGATCTGGGTGTACTTACCTGAGCTGTGATGACAAATATGGCGGTAGCGGAT 458  
 Db 181 CTTTTCGATCTGGGTGTACTTACCTGAGCTGTGATGACAAATATGGCGGTAGCGGAT 240  
 QY 459 GAATACGATGAGAGATGATGATCAAGTGAATTTTCCATGACCATGAGCTTCAAACT 518  
 Db 241 GAATACGATGAGAGATGATGATCAAGTGAATTTTCCATGACCATGAGCTTCAAACT 360  
 QY 519 GAAGCTACAAAGATATGAGTCTCCCTCCACACTTCCCTGTGATGAAACAGAACCACT 578  
 Db 301 GAAGCTACAAAGATATGAGTCTCCCTCCACACTTCCCTGTGATGAAACAGAACCACT 360  
 QY 579 GCCTATTTCCAGTCCAGCAGATTTGGTATAGAGAGGATTTGATGATGATGATGATGAT 638  
 Db 361 GGCATCTCTAGTCCAGCAGATTTGGTATAGAGAGGATTTGATGATGATGATGATGAT 420  
 QY 639 AGCTGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 698  
 Db 421 AGCTGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480  
 QY 699 ATCTTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 758  
 Db 481 ATCTTGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
 QY 759 GCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 818  
 Db 541 GCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600  
 QY 819 ATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 878  
 Db 601 ATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660  
 QY 879 CAGCGCCCTTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 938  
 Db 661 CAGCGCCCTTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 720  
 QY 939 GTTATTTCAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 998  
 Db 721 GTTATTTCAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 780  
 QY 999 TATTTCTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 1058  
 Db 781 TATTTCTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 840  
 QY 1059 CATTGATCAAGATGAGAAACTCACTGACAGAAATTTATCTAGCTATGACCTTAATTTGA 1118  
 Db 841 CATTGATCAAGATGAGAAACTCACTGACAGAAATTTATCTAGCTATGACCTTAATTTGA 900  
 QY 1119 TGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1178  
 Db 901 TGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960  
 QY 1179 CTTGAGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1238  
 Db 961 CTTTGAAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020  
 QY 1239 GAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1298  
 Db 1021 GAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080  
 QY 1299 GACATTTGAAGATGAGAGCGGAGAACTTTCAGCAGGAGCTGTGAGCTGCGAGAGCG 1358

Db	1081	GAAATTTGAAGATTAAGAAAGCGGGAGAACTTTCGAGCGAGGCGAACTCTGAGAGCTTGAGAAACGC	1140
Qy	1359	CCGCCAACGCCTCTTGGAGCAACGACCGCCAAAGACGAGAGCGGTTGGCTCACTGATGAGACG	1418
Db	1141	CAGGCAGGCTCTCTTGAGAGCAGACGCGCAAGGAGCGAGAGCGAGCTGGCTCACTGATGAGACG	1200
Qy	1419	CGCGAGCAGAGAGAGAAAGAGCGGAGCGCCCGAGAGCAGAGAGGCGCAAGCGCGCAGCTGGA	1478
Db	1201	CGCCGAGCAAGAGAGAAAGAGCGGAGCGCCCGAGAAACAGAGAGCCCGAAGAGCAGACTTGA	1260
Qy	1479	GCTGAGAGAGCAGCTGGAAGACGACGGGAGCTTGAGCGGCGAGCGAGAGAGAGAG	1538
Db	1261	GCTGGAGAAAGAGCTTAAGAAAAGACGCGGAGCTGAGAGAGGACAGAGAGAGAGAGAG	1320
Qy	1539	GAGGAGATCGAGAGGGCGGACGCGCAAAACGGGAACTGGAAGAGGCGAGCGCAACTTGA	1598
Db	1331	GAGGAGATCGAGAGGGCGGAGGAGCAGCAAAACGGGAGCTCGAACGAGAGAGAGAGAGAG	1380
Qy	1599	ATGGGAACGGAAACCGGAGACAGAACTCTGAATCAGAGGAAACAAAGACAGAGAGGCGAC	1658
Db	1381	ATGGGAACGGAAACCGGAGACAGAACTCTGAATCAGAGGAAACAAAGACAGAGAGGCGAT	1440
Qy	1659	CGTGGTCTTGAAAGCAGAGAGAAAGATCTTGAGAGTTGAGTTGAAGCTCTGATATGACAA	1718
Db	1441	CGTGGTCTTGAAAGCGAGAGAGAAAGATCTTGAGAGTTGAGTTGAAGGCTCTGATATGACAA	1500
Qy	1719	AAAGCATCAGCTAAGAAAGAAAACCTTCAGATATCAGGTTCAGCTGCGAACCCAGAGCA	1778
Db	1501	AAAGCATCAGCTGGAAGGAAACCTTCAGATATCAGGTTCAGCTGCGAACCCAGAGCA	1560
Qy	1779	AGAAATTGAGAGCAAGAACAAATCTTAGAGAGCTTAAGAAATGCTGAATATCAACCACTTACA	1838
Db	1561	AGAAATTGAGAGCAAGAACAAATCTTAGAGAGCTTAAGAAATGCTGAATATCAACCACTTACA	1620
Qy	1839	GACAGATTTGAGAGAAATCTCAGCAAAATGCTTGAGAAAGCTTATTCAGAGAAACAGATACT	1898
Db	1621	GCAACAGTTGAGAGAAATCTCAGCAGATGCTTGGAAGAACTTATTCAGAGAAACAGATACT	1680
Qy	1899	CAGTGACCACTTAAGAACAGTCCAGCAGAAACAGTTGCTATGAGACTGCTTCTTAACTCT	1958
Db	1681	CAGTGACCACTTAAGAACAGTCCAGCAGAAACAGTTGCTATGAGACTGCTTCTTAACTCT	1740
Qy	1959	CAAAAAGACCTTTGAGAACCAAGAGAGCTGGCGCGGACAGCTCCCGGAGACAGCTGGAAGCA	2018
Db	1741	CAAAAAGACCTTTGAGAACCAAGAGAGCTGGCGCGGACAGCTTCAGAGACAGCTGGAAGCA	1800
Qy	2019	GGTGAGAGAGAGACACAGGTCAAAGCTGACGAGAGATTGATGTTTTCAACCAACCACTGAA	2078
Db	1801	GGTGAGAGAGAGACACAGGTCAAAGCTGACGAGAGATTGATGTTTTCAACCAACCACTGAA	1860
Qy	2079	GGAATCTGAGAGAGATATCATAGCAAAACAGCACTTCAGAGACAGAGGTCCCTGAGGCGAC	2138
Db	1861	GGAATCTGAGAGAGATATCATAGTAACAGCACTTCAGAGACAGAGGTCCATGAGGCGAG	1920
Qy	2139	GCGACTTGAAGAGAAAGAGCAGAGAGAGAGAGACCTTGAGATTGAGAAACCAAAAGGAAGA	2198
Db	1921	GAGGCTTAAGAGAGAAAGAGCAGAGAGAGAGAGAGAGCTTGAGATTGAGAGAGCAAAAGGAAGA	1980
Qy	2199	CGCTCAGAGACGAGTTCAAGAAAGGAGCAAGCAATGCTTGAGACATGTGCACAGAGAGGA	2258
Db	1981	AGGTCAGAGAGAGATTTCAAGAAAGGAGCAAGCAATGCTGCAGAGACATGTCCACAGAGAGGA	2040
Qy	2259	GAGCAGACGCCCCCGGAAACCCCGACGAGAGAGACAGACTGGAAGAGGAGAGAGACAGTGTGAG	2318
Db	2041	GAGCAGACGCCCCCGGAAACCCCGACGAGAGAGACAACTGAAGAGGAGAGAGACAGTGTCAA	2100
Qy	2319	GAAAGAGAGCGGAGAGAGAGAGCCAGCCGAGAAATGCAAGCAACAGACAGTCCGCTTTT	2378
Db	2101	GAAAGAGAGCGGAGAGAGAGAGCCAGCCGAGAAATGCAAGCAACAGACAGTCCGCTTTT	2160
Qy	2379	CCATCCGATATGAGAGCCAGTAAAGCTGGCACCCAGAGACCTTGGTCTATCCACAGAGAA	2438
Db	2161	CCATCCATATTAAGAGCCAGTAAAGCCGG---GCGAGGACACCTTGGCCACAGAGAGAA	2217

Qy	2439	AGCCGCTTACCATTTCTGCAACAGAGATGTAAAGTGAATATTAACGAGCGCTGTA	2499
Db	2218	AGGTCCGCTTACAAATCTTCGCAAGAGATGTCCAAATGGTGTATTACGAGCGCTGTA	2277
Qy	2499	CCCCCTTGAATCCAGAAAGTCACGATGAGATCCATCCAGCCAGAGATATATGTCAT--	2555
Db	2278	CCCCCTTGAATCCAGAAAGTCATGACGAGATTCACATCCAGCCAGAGATATATGTCATG	2337
Qy	2556	-----GCTGATGAAAGCCAGATCGGACAGCCAGGATGGCTTGGAGGAGAGCT	2603
Db	2338	TAAAGGGAAATGGGTGATGAAAGCCAGCCAGGACGCGAGATGGCTTGGAGGAGAAC	2397
Qy	2604	GAAAGGAAACGCGATGGTTCCTCTGCAAACTATCGAAAAAGATTCCAGAAATGAGT	2663
Db	2398	GAAAGGAAACAGAGATGGTTCCTCTGCAAACTATCGAGAAAGATATCCAGAAATGAGAT	2457
Qy	2664	TCCCACTCCAGCCAAACACAGTACCGATTTGATCATCTGCCCCCTGCCCCAAACTGGCTCT	2723
Db	2458	TCCCACTCCAGCCAAACACAGTACCGATTTGATCATCTGCCCCCTGCCCCAAACTGGCTCT	2517
Qy	2724	GCGTGAGAACCCCTGCTCTTGGCAGTGAACCTTCTGAGCCCTTCACAAACCCCAACAA	2783
Db	2518	GCGTGAGAACCCCTGCTCTTGGCAGTGAACCTTCTGAGCCCTTCACAAACCCCAACAA	2577
Qy	2784	CTGGGCAGACTTCAGTTCACCGTGGCCAGACCTCAACGAGAAACCGAAACGGAACAA	2843
Db	2578	CTGGGCAGACTTCAGTTCACCGTGGCCAGACCTCAACGAGAAACCGAAACGGAACAA	2637
Qy	2844	CTGGGAAACGTGGGGGGGCTCAGCGCTTCTGTGACGCTACTAGTGTGGCCAGTTACGGCA	2903
Db	2638	CTGGGAAACGTGGGGGGGCTCAGCGCTTCTGTGACGCTACTAGTGTGGCCAGTTAAAGCA	2897
Qy	2904	GAGATCAGCCTTTTACCCCAAGCCACAGCCCACTGACCTCTCCCATCTCCGCTCTGGGCA	2963
Db	2698	GAGTGGGCTTCAACACAGCCACAGCCCACTGACCTCTCCCATCTCCGCTCTGGGCA	2757
Qy	2964	GGGTGAAAAGTGGAAAGGCTACAGCCGAACCTGTATCTCTGAGAGGCCAAAAAAGA	3023
Db	2758	GGGTGAAAAGTGGAAAGGCTCCACAGCAACACCTGTATCTCTGAGAGGCCAAAAAAGA	2817
Qy	3024	CAACCACTTAAATTTTAAACAAAAGTACGTCATCAACCGTTCTGGAACAGCAAGCATGTG	3083
Db	2818	CAACCACTTAAATTTTAAACAAAAGTATCTATCAACCGTTCTGGAACAGCAAGCATGTG	2877
Qy	3084	GTCGTTTGAAGAAAGTTCAAGGTCAGAAAGGTTGGTTCCTCCCAAGTCTTACGTGAATCTCAT	3143
Db	2878	GTCGTTTGAAGAAAGTTCAAGGTCAGAAAGGTTGGTTCCTCCCAAGTCTTACGTGAATCTCAT	2937
Qy	3144	TTTCAGGCCCCGTAAAGAAATCCACAAAGCATGATCTGACCTGACTGAAATCTCCGCTAG	3203
Db	2938	TTTCAGGCCCCGTAAAGAAATCCACAAAGCATGATCTGACCTGACTGAAAGCTCCCTCTAG	2997
Qy	3204	TCTTAAAGAGATGGCTTCCCGCGCCGACAGCCAGCCATTCCTCCGAGAAAGTTTATTC	3263
Db	2998	TCTTAAAGAGATGGCTTCCCGCGCCGACAGCCAGCCATTCCTCCGAGAAAGTTTATTC	3057
Qy	3264	CATGTACACATACGAGAGTTCTGAGCAGAGAAATTTAACTTTTACGCAAGGGAGTGTGAT	3323
Db	3058	TATGTACACTTAACGAGAGTTCTGAGCAGAGAAATTTAACTTTTACGCAAGGGAGTGTGAT	3117
Qy	3324	TGTGTTAACCAAGAAAGATGTGTACTGTGTGACAGGGAACGCTGGGGCGCAAGTCCGAGT	3383
Db	3118	TGTGTTAACCAAGAAAGATGTGTACTGTGTGACAGGGAACGCTGGGGCGCAAGTCCGAGT	3177
Qy	3384	CTTCCCTTCTAATATGTGAGGCTTAAAGATTCAAGAGGCTCTGGAACTGCTGAGAAAC	3443
Db	3178	CTTCCCTTCTAATATGTGAGGCTTAAAGACTCAAGAGGCTCTGGAACTGCTGAGGAAAC	3237
Qy	3444	AGGAGTTTAAAGAAAAAACCTGAATTTGGCCAGGTCATGTCTTCAACATGCTACGGG	3503
Db	3238	AGGAGTTTAAAGAAAAAACCTGAATTTGGCCAGGTCATGTCTTCAACATGCTACGGG	3297







Db 2413 TAAAGGGGAATGGGTGATGAAAGCCAACTGGAGAACCCGGCTGGCTTGAGAGAAATT 2472  
 Qy 2604 GAAAGGAGACGGGATGTTTCCCTGCAAACTATGAGAAAGAAATTCAGAAAGATAGGT 2663  
 Db 2473 AAAAGGAAAGACGGGGGTTCCTGCAAACTATGAGAGAAAGAAATTCAGAAAGATAGGT 2532  
 Qy 2664 TCCACCTCCGACCAAAACCAATGACCATCTGACATCTGCCCTCCCAAACTGGCTCT 2723  
 Db 2533 TCCCGCTCCGACCAAAACCAATGACCATCTGACATCTGCCCTCCCAAACTGGCTCT 2592  
 Qy 2724 GCGTGAGACCCCTGCTCTTTCAGTACCTTCTGAGCCCTCCCAAACTGGCTCT 2783  
 Db 2593 GCGTGAGACCCCGCCCTTTCAGTACCTTCTGAGCCCTCCCAAACTGGCTCT 2652  
 Qy 2784 CTGGGACAGACTTCACTTCCACGTCGCGCCAGCAGCTCAAAAGAGCCGAAACGAGCA 2843  
 Db 2653 CTGGGACAGACTTCACTTCCACGTCGCGCCAGCAGCTCAAAAGAGCCGAAACGAGTA 2712  
 Qy 2844 CTGGGATAGCTGGGCGGCTCAGCTTCTGACCGTACCTAGTCTGGCGCAGTTAGGCA 2903  
 Db 2713 CTGGGATAGCTGGGCGGCTCAGCTTCTGACCGTACCTAGTCTGGCGCAGTTAGGCA 2772  
 Qy 2904 GAGATGACCTTTTACCCAGCCACAGCAGCTGCTCCTCCCATCTCCCGTCTGGGCA 2963  
 Db 2773 GAGTCCGCTTTACTCTCAGCCACAGGCTGCTCCTCCCGTCTCTGTGTAGGCA 2832  
 Qy 2964 GGGTGAAGAGTGAAGGGCTACAGCGCAAGCCCTGTATCCTCGAGAGCCAAAAAGA 3023  
 Db 2833 GGGTGAAGAGTGAAGGGCTACAGCGCTATATCTTGGAGAGCCAAAAAGA 2892  
 Qy 3024 CAACCACTTAAATTTTAAACAAAGTATGTCATCACCGTCTGGAGCAACAAAGACATGTG 3083  
 Db 2893 CAACCACTTAAATTTTAAACAAAGTATGTCATCACCGTCTGGAGCAACAAAGACATGTG 2952  
 Qy 3084 GTGGTTGGAGAGTCAAGGTGAGAGGGTGTCCCAAGTCTTAAAGTAAATCAT 3143  
 Db 2953 GTGGTTGGAGAGTCAAGGTGAGAGGGTGTTCCTCCCAAGTCTTAAAGTAAATCAT 3012  
 Qy 3144 TTCAGGCGCGTGAAGAAATCCACAGATCGATCTGAGCCCTACTGAAAGTCTGCTAG 3203  
 Db 3013 TTCAGGCGCGTGAAGAGTCTACAGATCGATCTGAGCCCTACTGAAAGTCTGCTAG 3072  
 Qy 3204 TCTAAGAGAGTGGCTTCCCGCGCCCAAGCCAGCTTCCCGGAGAAAGTTATTC 3263  
 Db 3073 TCTAAGAGAGTGGCTTCCCGCGCCCAAGCCAGCTTCCCGGAGAAAGTTATTC 3132  
 Qy 3264 CATGTACACATACAGAGTCTGAGCAGAGATTTAACCTTTACAGAGGGGATGTGAT 3323  
 Db 3133 CATGTACACATACAGAGTCTGAGCAGAGATTTAACCTTTACAGAGGGGATGTGAT 3192  
 Qy 3324 TGTGTTTACCAAGAAAGATGTGACTGTGTGACGAGAAAGGTGGCGCAGATCCGAGT 3383  
 Db 3193 TGTGTTTACCAAGAAAGATGTGACTGTGTGACGAGAAAGGTGGCGCAGATCCGAGT 3252  
 Qy 3384 CTTCCTCTTCTAATATGTAGGCTTAAAGTTTACAGGGCTCTGCACTGCTGGGAAAC 3443  
 Db 3253 CTTCCTCTTCTAATATGTAGGCTTAAAGTTTACAGGGCTCTGCACTGCTGGGAAAC 3312  
 Qy 3444 AGGAGTTTAAAGAAAAAACTGAAATGCGCAGGTTATGTGCTCTAAGCTGACTAGG 3503  
 Db 3313 AGGAGTTTAAAGAAAAAACTGAAATGCGCAGGTTATGTGCTCTAAGCTGACTAGG 3372  
 Qy 3504 TCCGGAACAATCACTGCTGCTCTGGGAGCTGATCTTGTATCCGAAAAAGAACCCAG 3563  
 Db 3373 TCCGGAACAATCACTGCTGCTGCTGGGAGCTGATCTTGTATCCGAAAAAGAACCCAG 3432  
 Qy 3564 TGGATGTGGAGAGGAACTGCACTGAGGAGAAAGCCGAGATGGGTGTTTC 3623  
 Db 3433 TGGATGTGGAGAGGAGCTGCACTGAGGAGAAAGCCGAGATGGGTGTTTC 3492  
 Qy 3624 AGCAAAATTTATGCAAACTTAAAGCCCGGAAACAAAGCAAAATCAACCCAGAGTACC 3683  
 Db 3493 AGCAAAATTTATGCAAACTTAAAGCCCGGAGAGCAAAATCACTCAACAGAGCACCC 3552

Qy 3684 CAAGCCGAGTGCAGCCAGAGTGTGCGCAGTGTATCGGATGATGACATTAACCCGCCA 3743  
 Db 3553 TAACTCAACAGCATTTAGCGGAGTGTGCGCAGTGTATGAGATGACACCCGCCA 3612  
 Qy 3744 GACGATGACGACATGACCTTACGAAAGGCAATATCAACTCTCTCAACAGAGAGA 3803  
 Db 3613 GACGATGACGATGAGCTGTGCTTCAACAGAGGCAATATCAACTCTCTCAACAGAGAGA 3672  
 Qy 3804 CCGGAGCTGTGAGAAAGAGAGTCACTGGGCAAGTGGGCTCTTCCATCAATATATGT 3863  
 Db 3673 CCGTACCTGTGTGAGAAAGAGAGTCACTGGGCAAGTGGGCTCTTCCATCAATATATGT 3732  
 Qy 3864 AAAGCTGACCAAGACATGAGCCCGAGCAGCAATGATCATATGTTGTCATCCCCC 3923  
 Db 3733 GAAGCTGACCAAGACATGAGCCCGAGCAGCAATGATCATATGTTGTCATCCCCC 3792  
 Qy 3924 TCAGGCTTGAAGTCTTCAAGAGACCCACTATCCATATCACTGCCCCAGAGGATATG 3993  
 Db 3793 TCAGGCTTGAAGTCTTCAAGAGACCCACTATCCATATCACTGCCCCAGAGGATATG 3852  
 Qy 3984 GAGATGAGCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4043  
 Db 3853 GAGATGAGCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3912  
 Qy 4044 GAATCTGACGAGACAGTATTACCTCATTTGACCTTATGATGATGATGATGATGATG 4103  
 Db 3913 GAATCTGACGAGACAGTATTACCTCATTTGACCTTATGATGATGATGATGATGATG 3972  
 Qy 4104 AGTCACTGCGTGCAGAGGAGAGACA--MATGAGAACTGCAGAGGTGGTGGTCTT 4161  
 Db 3973 AGTATTAATCTGCAGAGATAGAGCAAAATTTACAAAACACAGAGGTGGTGGTCTT 4032  
 Qy 4162 TTGGGCTTTTCTTATGACCTGACCTGAC--CGGCGCGCTTCAAGAGGGGCTTTCAT 4220  
 Db 4033 TTGGGCTTTTCTTATGACCTGACCTGACCTTTCCTTTCATTAATTAATTAATTAAT 4092  
 Qy 4221 A-GTTTAAAGATTTTAAATGATTTTAAATGATTTTAAATGATTTTAAATGATTTTAA 4279  
 Db 4093 AGTATTAATTTTAAATGATTTTAAATGATTTTAAATGATTTTAAATGATTTTAAAT 4152  
 Qy 4280 ACTTCTTGGCTATTTTGGTTTACAAAAACCCACTATCAGAGAGTGGTGGTCTT 4339  
 Db 4153 GACTTCTTGGCTATTTTGGTTTACAAAAACCCACTATCAGAGAGTGGTGGTCTT 4207  
 Qy 4340 ACGATTAATTAATGCTGTTCCGGCGCTACGATTAAGTGAAGTGGTGGTGGTGGT 4399  
 Db 4208 GTGCTATTTAAATTTTCCAAATGTCATTAATCTGAGACTGATGATTTTTCATTT 4267  
 Qy 4400 TGTCCAGTGTCCCAACACATTTGTGATTTTGGGCTGTTCC--TGCCTGAGACA 4455  
 Db 4268 TGTCCAGTGTACCAACATTAATTTGTGAGTTTGGGCTTTTCCCTTACATGAGAGTG 4327  
 Qy 4456 CAGAGAGATGGGTGATCTGTTTGAATTTGTATGATGATGATGATGATGATGATG 4515  
 Db 4328 CAGAGAGATGGGTGATCTGTTTGAATTTGTATGATGATGATGATGATGATGATG 4386  
 Qy 4516 GGGGTTATGCTTGTCTGACCAATCACTGATC--TGTGCGAGATGATGATGATGATG 4574  
 Db 4387 GTGTTTGTGTTTGTGTTGTGATGATGATGATGATGATGATGATGATGATGATG 4446  
 Qy 4575 AAGAGATGAT--CTTCTGATGATGATGATGATGATGATGATGATGATGATGATG 4633  
 Db 4447 CATTAAGAAATGATTTCTTTCATGAGGAAAGTATTTACTTGTGATGATGATGATG 4506  
 Qy 4634 --GCATTCATTTTCTTTCATGAGTACCTTTGATGATGATGATGATGATGATG 4690  
 Db 4507 TGCATTTAATTTTATTTTGAACAGTACCTTTGATGATGATGATGATGATGATG 4566  
 Qy 4691 CCGTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4750  
 Db 4567 TTGTTGCTGCTCAATTTATCTGTTGATGATGATGATGATGATGATGATGATGATG 4626

QY 4751 ATGCTTCACACCATTAAGTAAAGCCCACTGCCAGGCGTTACGATCATAGTA 4810  
 Db 4627 GGTGTTGATAGATTAATACAGTACAGAAC-ACCAGTGAAGCGTTAGATCAAGATA 4685  
 QY 4811 CCACACGTC-----TTAGTCTCTGTTACGTTGAAG-TTTATCCAGTTGCTTTTATGGA 4864  
 Db 4686 TCACAGCTCTCTTTTGTCTGTTACATGAAGTTTATCCAGTTACTTTCATGGGA 4745  
 QY 4865 -TATCTGAACAGTAACTCTCTTGACAGAAAGATGATAGAGTCTCCCTGCA 4919  
 Db 4746 TGACCATTTTGAACAAGTAAATTTCTTGACAAAGAAAGATGATAGAGTCTCCCTGCA 4805  
 QY 4920 ATTAATTTCCAGTGTTCATTTTTCATTAACCTGCTGGGGTCTCTACATTAAT 4979  
 Db 4806 ATTAATTTT-CCAAATGTTTAACTTTTAACTAGACTGT-GGAATTTCTACAGATTAAT 4863  
 QY 4980 GAATGCGGCTCGTGGCGCTGTGTGTAACTGTGGCTGTAGTGAAGCCGTGTGT- 5037  
 Db 4864 GAATGAGCTCATGTGCTGCTTTGTGTGA-AAATGCTGTAGTGAAGCCGTGTGT 4922  
 QY 5038 -CCTAGATATTAGTGAAGTGGGAAGGAATT 5071  
 Db 4923 CTTTAAACACTAGTGAAGCTCTCAATAAAT 4957  
 RESULT 6  
 AF132672 3812 bp mRNA linear ROD 28-JUN-1999  
 AF132672  
 LOCUS Rattus norvegicus EH-domain/SH3-domain containing protein mRNA,  
 complete cds.  
 ACCESSION AF132672  
 VERSION AF132672.1 GI:4838525  
 KEYWORDS  
 SOURCE Rattus norvegicus.  
 ORGANISM Rattus norvegicus.  
 Rattus norvegicus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 REFERENCE 1 (bases 1 to 3812)  
 AUTHORS Okamoto, M., Schoch, S. and Sudhof, T. C.  
 TITLE BSH3/intersectin, a protein that contains EH and SH3 domains and  
 binds to dynamin and SNAP-25. A protein connection between  
 exocytosis and endocytosis?  
 J. Biol. Chem. 274 (26), 18446-18454 (1999)  
 MEDLINE 99303609  
 PUBMED 10373452  
 REFERENCE 2 (bases 1 to 3812)  
 AUTHORS Okamoto, M., Schoch, S. and Sudhof, T. C.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-MAR-1999) Center for Basic Neuroscience and HHMI,  
 Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX  
 75235, USA  
 FEATURES  
 source  
 1. 3812  
 /organism="Rattus norvegicus"  
 /db\_xref="taxon:10116"  
 41. 3481  
 /note="BESH1"  
 /codon\_start=1  
 /product="EH-domain/SH3-domain containing protein"  
 /protein\_id="AD31026.1"  
 /db\_xref="GI:4838526"  
 /translation="MAQFPFGSLDIWATVEERAKDOFOSLKPISGFTGDOA  
 RNFPGSGPVLAIOWALADNMKDMQVFSIAMLKILKLGQYOLPALPVM  
 KOOPAISSAPAGCIGMAGMPLTAAPVPMGSIIVMGSPVSVAPAAVPLAN  
 GAPVIOPLPAHPAATLTPKSSFSRSGSGLNTKLOKQSFVYASPPAAEMAVP  
 QSSRLKRLQFNHDKTMSGLTGPQARTILWQSSLPQOLASIMWLSIDQGLKTA  
 EEPILAMHLIDVAMSGQPLPVLPEVTPSPFRVSGSGSVISSSASDQPLPEPS  
 SEDQOYEKLPVTFEDKRENERGNLEKREKQALLCOQREKRELAQLBARER  
 KERBROEKRRKLELEKOLEKORELEBERREKLEIRERAKKELROROLEMER  
 NRROELTQNKQOEGIVTKARKTLFEFLALNKKHLLKGLKIDICRLATROE  
 IESTNKRRLRIALITHLQOOLDSQMLKRLIPKQLISDQIKQYQNSLRDSLT  
 LKRALEKELARQOLREOLDEVERETSKLDELIDVNNOLKELREIHSKQOLOKRSI

BASE COUNT 1088 a 972 c 1058 g 694 t  
 ORIGIN  
 Query Match 61.0%; Score 3103.6; DB 10; Length 3812;  
 Best Local Similarity 88.4%; Pred. No. 0;  
 Matches 3535; Conservative 0; Mismatches 229; Indels 234; Gaps 6;  
 QY 219 GAGATGAGGCGTCGATCAGAGGTAACGTAAATAGAACCATGGCTCAATTTCCACACC 278  
 Db 1 GAGATGAGGCGTCGATCAGAGGTAACGTAAATAGAACCATGGCTCAATTTCCACACC 60  
 QY 279 TTTGCGTGTAGCTGATGTCTGGGCTATTACTGTGGAGAAAGGCGCAAGATGACCA 338  
 Db 61 TTTGCGTGTAGCTGATGTCTGGGCTATTACTGTGGAGAAAGGCGCAAGATGACCA 120  
 QY 339 GCGTTCCTTAGCTCGAAGCGGATAGCGGATTTATTAAGTGTGATCAAGCGAGAACTT 398  
 Db 121 GCGTTCCTTAGCTCGAAGCGGATAGCGGATTTATTAAGTGTGATCAAGCGAGAACTT 180  
 QY 399 TTTTTCGAATCTGGGTTACCTGACCTGTCTTACCAAAATATGGGCGCTAGCGACAT 458  
 Db 181 CTTTTCGAATCTGGGTTACCTGACCTGTCTTACCAAAATATGGGCGCTAGCGACAT 240  
 QY 459 GAATTAACATGGAAGATGATGATCAAGTGAATTTTCATAGCATGAAGCTTCAACT 518  
 Db 241 GAATTAACATGGAAGATGATGATGATCAAGTGAATTTTCATAGCATGAAGCTTCAACT 300  
 QY 519 GAAGTCAAGATATATAGCTCCCTCCACACTTCCCTGTCTATGAAGACGACCAAGT 578  
 Db 301 GAAGTCAAGATATATAGCTCCCTCCACACTTCCCTGTCTATGAAGACGACCAAGT 360  
 QY 579 GGCTATTTCCAGTGCACCAAGATTTGGTATAGAGGATTTGTCATGTCACACACTCAC 638  
 Db 361 GGCTATTTCCAGTGCACCAAGATTTGGTATAGAGGATTTGTCATGTCACACACTCAC 420  
 QY 639 AGCTGTGCTCCTGCGCAATGGGCTCATTCAGTTGTGAATGTCTCCACCCTTAGT 698  
 Db 421 AGCTGTGCTCCTGCGCAATGGGCTCATTCAGTTGTGAATGTCTCCACCCTTAGT 480  
 QY 699 ATCTTCTGCTCCAGCAGAGGATGCTCCCTGCTTACGAGGAGCTCTCCCGTCATACA 758  
 Db 481 ATCTTCTGCTCCAGCAGAGGATGCTCCCTGCTTACGAGGAGCTCTCCCGTCATACA 540  
 QY 759 GGCTTCTGCTCCTGCGCAATGGGCTCATTCAGTTGTGAATGTCTCCACCCTTAGT 818  
 Db 541 GGCTTCTGCTCCTGCGCAATGGGCTCATTCAGTTGTGAATGTCTCCACCCTTAGT 600  
 QY 819 ATCTGATCAGGATCAGATTAACACTTAAGTTACAGAGGACCAATCATTTGATGTGC 878  
 Db 601 ATCTGATCAGGATCAGATTAACACTTAAGTTACAGAGGACCAATCATTTGATGTGC 660  
 QY 879 CAGGCGCCCTCCAGCAGAGGATGAGCTGCTCCCTCAATCAAGGCTGGAATACAGGCA 938  
 Db 661 CAGGCGCCCTCCAGCAGAGGATGAGCTGCTCCCTCAATCAAGGCTGGAATACAGGCA 720  
 QY 939 GTTATTAACAGCAGCAAAATCATAGTGAACATTAACAGGCTCCCGCAGCAAGAC 998  
 Db 721 GTTATTAACAGCAGCAAAATCATAGTGAACATTAACAGGCTCCCGCAGCAAGAC 780  
 QY 999 TATTCTCATGCAATCAAGTTTACCCAGGCTAGCTGCTTCAATATGAAATCTTTTCTGA 1058  
 Db 781 TATTCTCATGCAATCAAGTTTACCCAGGCTAGCTGCTTCAATATGAAATCTTTTCTGA 840



QY	1059	CATTGATCAAGATGAGAAAATC	CACTCAGACAGAAATTAATC	CTAGCTATGACACTAAATTGA	1118
Db	841	CATTGATCAAGATGAGAAAGCT	CAACCGACAGAAATTCATCT	TAAGGATGCACCTGATTGA	900
QY	1119	TGTTGCCATGTCTGTGTCAAG	CACTGCGCCGCTCTGCTCC	AGAAATAATCCTCCTTC	1178
Db	901	TGTTGCTATGTCTGTGTCAAG	CACTGCGCCGCTGTCTGCT	CCAGAAATCATCTCCCTTC	960
QY	1179	CTTGAGAAAGCTTGCTCCGG	CACTGGATGTTCCTCAT	AACTCTTCTTCTGTGTGATCA	1238
Db	961	CTTTAGAAAGAGTTTGCTCTT	CGCAGTGGGATGTTCCTCAT	AACTCTTCTTGTGTGAGACCA	1020
QY	1239	GAGGCTCCTCGAAGAGCCGTG	CAAGGGATGTGACGACGCC	AGAAAGAACTGCGCTGT	1298
Db	1021	GGGGCTCGCCGAGAGACCAT	GTCTGACAGATGTCTCCGAT	CACTAACTCTTGTGTGAGACCA	1080
QY	1299	GACATTTGAAGATTAAGAAAG	CGGGAGAACTTTCAG	CCAGGCACTGTGTGAGCTTG	1358
Db	1081	GACATTTGAAGATTAAGAAAG	CGGGAGAACTTTCAG	CCAGGCACTGTGTGAGAAACG	1140
QY	1359	CCGCCAAGCCGCTCTTGAG	ACGACAGCGCAAAAGCAG	AGCCGTTTGCTCACTGTGAGCG	1418
Db	1141	CAGCGAGGCCCTCTCTGAG	CAAGCAGCGCAAAAGGAG	CGCGCTGTGCTCACTGTGAGCG	1200
QY	1419	CGCCGACGAGAGAGGAAAG	AGCGGGGACGCCCGAC	GCAGAGAGAGGCCCAACCG	1478
Db	1201	CGCCGACGACAGAGAGGAA	AGCGGGGACGCCCGAC	GCAGAGAGAGGCCCAACCACTGA	1260
QY	1479	GCTGTGAGAAAGCAGCTTG	AGAAAGACGCGGAGCTGT	GCAGCGCAGCGAGAGAGAGAG	1538
Db	1261	GCTGTGAGAAAGCAGCTTA	BAAGAAAGCGGGAGCTGT	GCAGAGAGCAGAGAGAGAGAG	1320
QY	1539	GAAAGAGATCGAGAGCGCG	ACGCGCAAAACGGGAACT	GTGAAAGGCAAGCAACTTTGA	1598
Db	1331	GAAAGAGATCGAGAGCGCG	ACGCGCAAAACGGGAGCT	GTGAAACGGCACTGCTTGA	1380
QY	1599	ATTGGAAACGGAACCGG	AGACAGGAACCTCGAT	TCAGAGGAACAAGAGGACAG	1658
Db	1381	ATTGGAAACGGAACCGG	AGACAGGAACCTCGAT	TCAGAGGAACAAGAGGAGGAG	1440
QY	1659	CGTGTCTCTGAAGCAGAG	AGAAAGACTGTGAGTTG	AGACTCTTAATGACAA	1718
Db	1441	CGTGTCTCTGAAGCAGAG	AGAAAGACTGTGAGTTG	AGACTCTTAATGACAA	1500
QY	1719	AAAGCATCAGCTAGAG	AAAACTTCAGATATC	AGGTGTGACTTGGCAAC	1778
Db	1501	AAAGCATCAGCTAGAG	AAAACTTCAGATATC	AGGTGTGACTTGGCAAC	1560
QY	1779	AGAAATTGAGAGCA	CGAAACAGTCTTA	AGAGGCTTAAGAAATG	1838
Db	1561	AGAAATTGAGAGCA	CGAAACAGTCTTA	AGAGGCTTAAGAAATG	1620
QY	1839	GCAGAGTTTGAGAGAT	CTCTGCAAAAGCTT	GTGAAAGCTTAATCC	1898
Db	1621	GCAGAGTTTGAGAGAT	CTCTGCAAAAGCTT	GTGAAAGCTTAATCC	1680
QY	1899	CAGTGACCACTTAA	AAACAGTCCAGCA	GAACAGTTTGACAT	1958
Db	1681	CAGTGACCACTTAA	AAACAGTCCAGCA	GAACAGTTTGACAT	1740
QY	1959	CAAAAGACCTTTG	AGACAAAGAGAGCT	GGCCGCGACAGCT	2018
Db	1741	CAAAAGACCTTTG	AGACAAAGAGAGAGCT	GGCCGCGACAGCT	1800
QY	2019	GGTGAGAGAGAGAG	CAAGAGTCA	AAAGCTGTGAGAGAT	2078
Db	1801	GGTGAGAGAGAGAGAG	CAAGAGTCA	AAAGCTGTGAGAGAT	1866
QY	2079	GGAACTGAGAGAT	CACTATGACAAAC	CACTCCAGAGCAGAGG	2138
Db	1861	GGAACTGAGAGAT	CACTATGACAAAC	CACTCCAGAGCAGAGG	1922
QY	2139	GGCACTGAAGCA	GAAGAGCGAGAGAGAG	AGCCTTGAAGT	2196

Db	1921	GAGGCTTAAGCAAAAGACGACGAGAGGAAAGAGCCTGGAGTTTGAGAAAGCAAAAGGAAGA	1980
Qy	2199	CGCTCAGAGACGAGTTTCAGGAAGGGACAAACGAATGSGTGGAGCATGTGCAGCAGGAGGA	2258
Db	1981	AGGTCAGAGACGAGTTTCAGAAAAGGACAAAGCAATGCGAGGAGCATGTTCACGACAGAGGA	2040
Qy	2259	GCAGCCACGCGCCCGGAAACCCACGAGGAGGACGACCTGAAAGAGGGAAAGCACTGTCTAG	2318
Db	2041	GCAGAGAGCCCTCTGGAAACCCCAAGAGGAGCAAACTGAAAAGGGAAGCACTGTCTAA	2100
Qy	2219	GAGGAGGAGCGCGGAAAGAGAGAGCCCAAGCCGGAAATGCAAGCAAGCAGAGTCCGCTTTT	2378
Db	2101	GAGGAGGAGCGCGGAAAGAGAGAGCCCAAGCCGGAAATGCAAGCAAGCAGAGTCCGCTTTT	2160
Qy	2239	CGATCCGCAATCAGAGACGAGCTAAGTGGCCACCCAGGACCCCTGATACCAAGAGAA	2438
Db	2161	CGATCCCAATCAAGAGAGCCAGCTAAGCCGG---CGAGGACCCCTGGCCCAAGACAGAGA	2217
Qy	2439	AGGCGCCGCTTACCATTTCTGCACAGAGAGTGTAAAGTGTATTTACGAGCGCTGT	2498
Db	2218	AGTCCGCTTACCAATCTCTGCACAGGAGAGTGCCTCAAGTGTGTATTTCGAGCGCTGT	2277
Qy	2499	CCCCCTTGAATCCAGAAAGTCAGATAGATATCAACATCCAGCCAGGAGATATAGCAT---	2555
Db	2278	CCCCCTTGAATCCAGAAAGTCAAGATATCAACATCCAGCCAGGAGATATATGATAGT	2337
Qy	2556	-----GCTGATGTAAAGCCAGCATGAGAGCCAGAGATGCTTGGAGAGAGCT	2603
Db	2338	TAAAGGGGAATGGGTGGATGTAAAGCCAGACCGGAGGCCAGAGATGCTTGGAGAGAAC	2397
Qy	2604	GAAAGGGAAGACGGAGATGTTCCCTGCAAACTATGCAGAAAAAGATTCCAGAAAAATGAGT	2663
Db	2398	GAAAGGGAAGACAGGATGTTCCCTGCAAACTATGCAGAAAAAGATTCCAGAAAAATGAGAT	2457
Qy	2664	TCCACATCCAGCCAAACCAAGTGCACGATCTGACATCTGCGCCCTGCGCCCAAACTGGAGCT	2723
Db	2458	TCCACATCCAGCCAAACCAAGTGCACGATCTGACATCTGCGCCCTGCGCCCAAACTGGAGCT	2517
Qy	2724	GCGTGAGACCCCTGCTCTCTTGTGCAGTGAACCTTCTTGAGCCCTTCCACAAACCCCAACA	2783
Db	2518	GCGTGAGACCCCTGCTCTCTTGTGCAGTGAACCTTCTTGAGCCCTTCCACAAACCCCAACA	2577
Qy	2784	CTGGGCGAGCTTCAGTTCCACGTCGGGCCAGCGAGCTCAACGGAAGCCAGAAACGGACA	2843
Db	2578	CTGGGCGAGCTTCAGTTCCACGTCGGGCCAGCGAGCTCAACGGAAGCCAGAAACGGACA	2637
Qy	2844	CTGGGATACGTCGGCGGCTCAGGCTTCTCTGACCGTACCTAATGTCTGGCCAGTTACGGCA	2903
Db	2638	CTGGGACACGTCGGCGGCTCAGGCTTCTCTGACCTGAATCCCAATGCTGGCAGTTAAGGCA	2697
Qy	2904	GAGATCAGCCTTTTACCCAGCCACAGCCACTGTGCTCTTCCCATCTCCGCTCTGGGCCA	2963
Db	2698	GAGGTCGGCTTCTCAACACAGCCACAGCCACTGTGCTCTTCCCATCTCCGCTCTGGGCCA	2757
Qy	2964	GAGTGAAAAAGTGGAAAGGCTTACAGCGGAAGCCGTATCTCCTGAGAGCCAAAAAAAGA	3023
Db	2758	GAGTGAAAAAGTGGAAAGGCTTCAAGCAGACAGCCGTATCTCCTGAGAGCCAAAAAAAGA	2817
Qy	3024	CAACCACTTAAATTTTACAAAAAGTACGTATCACCGTTCTGGAAACAGCAAGACATGTG	3083
Db	2818	CAACCACTTAAATTTTACAAAAAGTATGATCATCACCGTTCTGGAAACAGCAAGACATGTG	2877
Qy	3084	GTCGTTTGGAGAAATTCGAAGTCAGAAAGGTTGGTTCCCAAGTCTTACGTGAAACTCAT	3143
Db	2878	GTCGTTTGGAGAAATTCGAAGTCAGAAAGGTTGGTTCCCAAGTCTTACGTGAAACTCAT	2937
Qy	3144	TTTCAGGCGCCCTTAAGAAATCCAAAGCATTCGATCTGCGCCCTTACGTAAAGTCTGCTAG	3203
Db	2938	TTTCAGGCGCCCTTAAGAAATCCAAAGCATTCGATCTGCGCCCTTACGTAAAGTCTGCTAG	2997
Qy	3204	TTTAAAGAGATGCTTCCCCCGCCGCAAGCCAGCATTTCCCGAGAAAGATTATTC	3263





QY	2223	GGACAGCAAAAGGCTGGGAGCATGCGACGAGAG---	3AGCAGCCACGCGCCCGGAAAC	2279
Db	2236	GGACAGCATGCGCTGGAGCATGTGCGACGAGGACGAGCATCAGACGCCAAGAAACT	2295	
QY	2280	CCACGAGGAGCAGACTGGAAGAGGGAAGCACTGTCCAGAAAGAGAGCGGAAAGAG	2339	
Db	2286	CCACGAAAGAGAAAACTGAAAAAGGAGGAGAGTGTCTAAAAAGAAAGATGCGAGAA	2355	
QY	2340	AGCCAAAGCCGGAATTCGAAGCAAGCAAGATCGGCTTTTCCATCCGCGATCAGAACCGAC	2359	
Db	2356	AGGCAAAAGGAAACACAAAGCAAGCTGGGTGGCTTTTCCATCAACACAAAGAACGAC	2415	
QY	2400	TAAGCTGGCCACCCGAGGACCCCTGGTCTACACAGAGAAAGGCCGCTTACCATTTTCG	2459	
Db	2416	TAAGCAGCTGTCCAGGAGCCCTGTGTCACTGCAGAAAAAGGTCCACTTACCATTTTCG	2475	
QY	2460	ACAGAGAGGTAAAAAGTGGTATTTACGAGCGCTGTACCCCTTTGAATCCAGAAATCA	2519	
Db	2476	ACAGGAAAAATTTAAAGTGGTGTATTCGGGAGACGTACCCCTTTGAATCCAGAAAGCA	2535	
QY	2520	CGATGAGATCCACATCCAGGACGAGAGATATAGTCAT-----GGTGGATGA	2564	
Db	2536	TGATGAAATTCATTCAGCCAGAGACATAGTCATGTGTTAAAGGGGAATGGGTGGATGA	2595	
QY	2565	AAGCCAGCTGAGAGCCAGAGTGGCTTGGAGAGAGCTGAAAAGGAGACGGGATGGTT	2624	
Db	2596	AAGCCAACTGGAAACCCGGCTGGCTTGGAGAGAAATTTAAAGGAAACAGGGTGGTT	2655	
QY	2625	CCCTGCAAACTATCGAGAAAAATTCAGAAAAATGAGTTCCCACTCCAGCCAAACCACT	2684	
Db	2656	CCCTGCAAACTATGCAAGAAAAATCCAGAAAAATGAGTTCCCGCTCCAGTGAAACCACT	2715	
QY	2685	GACCGATCTGACATCTGCCCCCTGCCCCCAACTGGCTCTGCTGAGACCCCTGCTCTT	2744	
Db	2716	GACTGATTCACATCTGCCCCCTGCCCCCAACTGGCTCTGCTGAGAACCCCGCCCTTT	2775	
QY	2745	GCCAGTAGCCCTTCTGTAGCCCTCCACAACCCCAACACTGGGCGAGTTCAAGTTCCAC	2804	
Db	2776	GGCGATTAACCTTCTTGAGGCCCTCCAGACCCCTTAATTACTGGGCCGACTTCAAGCTCAC	2835	
QY	2805	GTGGCCCGACGCTCAACAGAAAGCCAGAAACGCAACTGGGATACCTGGCGGCTCA	2864	
Db	2836	GTGGCCCGACGACAGATGAAAGCCAGAAACGGAATTACTGGGATGTCATGGGACGCCA	2895	
QY	2865	GCCCTTCTGACCGTACTAGTGTGTGGCCGCTTACGGCAGAGATCAGCCTTTACCCGAC	2924	
Db	2896	GCCCTTCTTCCCGTTCCTCAAGTGGCGGCGAGTTAAGCAGAGGTCCGCCCTTTACCTCAGC	2955	
QY	2925	CACAGCCACTGGCTCTGCCCATCTCCGCTCTGGGCGCAGGAGTGAAGAGTGGGAAGGCT	2984	
Db	2956	CACGGCCACTGGCTCTGCCCATCTCCGCTCTCTGTGTGTCAGGCGCAGGAGTGAAGAGGCGCT	3015	
QY	2985	ACAAAGCGAAGCCCTGTATCCCTGAGAGCCAAAAAGACAAACACTTAATTTTAAACA	3044	
Db	3016	ACAAAGCTCAAGCCCTTAATCTTGTGAAGGCCAAAAAGACAAACACTTAATTTTAAACA	3075	
QY	3045	AAGGAGCGCTATCCCGTCTGGAACAGCAAGCAATGTGGTGGTTTGGGAAGTTCAAG	3104	
Db	3076	AAATGATGTCTATCCCGTCTGTGAACAGCAAGCAATGTGGTGGTTTGGGAAGTTCAAG	3135	
QY	3105	TCAGAAAGGTTGTTCGCCCAAGTTTACGTGAACCTCATTTCCAGGGCCGTTAAGGAAATC	3164	
Db	3136	TCAGAAAGGTTGTTCGCCCAAGTTTACGTGAACCTCATTTCCAGGGCCCATTAAGGAATC	3195	
QY	3165	CACAAAGCATTCATCTGGGCCCTACTGAAGTCTGTCTAAGTTTAAAGAGTGGCTTCCC	3224	
Db	3196	TACAAAGCATTCATCTGGTCTTTCAGAGATCTGTCTAAGTTTAAAGCGATGCTCTCC	3255	
QY	3225	GGCGCCCAAGCACCACTATCCCGGAGAAAGTTTATTCGCATCTACACTAAGAGATTC	3284	
Db	3256	AGCAGCCAAAGCGGTGCTTCGGGAGAAAGATTTATTCGCATCTAAGACTTAAGAGATTC	3315	
QY	3285	TGAGCAAGAGATTTAACTTTTACGCTTTCAGCAAGGGAGTGTGATTTGTGGTTACCAAGAAAGATGG	3344	

Db	3316	TGAGCAAGAGATTTAACCTTTCAGCAAGGGAGTGTATTTTGTTTACCAAGAAAGATTG	3375
QY	3345	TGACTGTGTGACGGGAACGGTGGGCGACAAGTCCGGAGTCTTCCCTTAACTATGTAG	3404
Db	3376	TGACTGTGTGACAGGAACAGTGGGCGACAAGCCGGAATCTTCCCTTAACTATGTAG	3435
QY	3405	GCTTAAAGATTACAGAGGCTCTGGAACTGTGGGAAAAACAGAGATTAGAAAAAAACC	3464
Db	3436	GCTTAAAGATTACAGAGGCTCTGGAACTGTGGGAAAAACAGAGATTAGAAAAAAACC	3495
QY	3465	TGAAATTGCCAGGTATTGTCTTCTTACGCTGCTACTGTGTCCTCCGAACACTACCTTGGC	3524
Db	3496	TGAATTGCCAGGTATTGTCTTCTTACCAACCGCCACCGGCCCGGACGCTCAGCTTCGC	3555
QY	3525	TCTCGGGACGTGATTTCTGATCCGGAAAAAACAACCAAGTGGATGTGGAAAGGAAGT	3584
Db	3556	CCCTGTGACGTGATTTTGTATCCGAAAAAGAACCCAGTGGATGTGTGGAAAGGAGACT	3615
QY	3585	GCAACCTCGAGGGAAAAAGCCGCAAGATGAGGGGTGTTTCAGCAAAATTATGTCAAATTCT	3644
Db	3616	GCAACCACTGTGGAAAAAGCCGCAAGATGAGCTGTGTTCCAGCTAATTATGTAAAGCTTCT	3675
QY	3645	AAGCCCGGAACAGCAAAATATCAACCCCACTGAGTACCCAAAGCCGAGTGCAGCCAGC	3704
Db	3676	AAGCCCTGGAGCAGCAAAATATCACTCAACAGAGCACTTAAGTCAACAGATTAAGCCGC	3735
QY	3705	AGTGTCCAGGTGATTCGGAGTACGATTTACACCGCCCAAGACGATGACGAATAGCTT	3764
Db	3736	AGTGTCCAGGTGATTTGGAGTGTAGACTTACACCGCCCAAGATGACGATGAGCTGGCTT	3795
QY	3765	CAGCAAGCCAGATCATCAACGTCCTTAACAGAGAGACCCGAGCTGTGGAAAGAGGA	3824
Db	3796	CAACAAAGGCCAGATCAATCAACGTCCTTAACAGAGAGACCCCTGATGTGGTGAAGAGGA	3855
QY	3825	AGTCAGTGGGCAAGTGGGCTCTTCCCATTCATTTATTAAGCTGACACACATGGA	3884
Db	3856	AGTCATGTGACAAAGTGGGCTCTTCCCATTCATTTATGTGAAGCTGACACACATGGA	3915
QY	3885	CCCAAGCCAGCAATG 3899	
Db	3916	CCCAAGCCAGCAATG 3930	
RESULT 8			
AF064244			
LOCUS			
DEFINITION	AF064244	7247 bp	mRNA
ACCESSION	AF064244		linear
VERSION	AF064244.1		PRI 21-NOV-1998
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			



Qy	1362	CGAAGCCGCTTTGGAGCGACGCGCAAAAGAGAGAGACGGTTTGGCTCACTGTGACCGCGC	1423
Db	1213	GCAAGCTCTCTGGAAACAGACCGCCCAAGAGACAGAGACCCCTTGGCCAGCTGGACGGGC	1272
Qy	1422	CGAGCAGAGAGGAAAGACCGGAGCGCCGAGAGCAGAGAGGCCAAGCGGACGTGGAGCT	1481
Db	1273	GGAGCAGAGAGAGAAAGAGCGCTTGAGCCGCCAGAGCAAGCCGCAAAAGCAACTGGAACT	1332
Qy	1482	GGAGAGAGAGCTGGAGAAAGCAGCGGGGCTGGAGCGGCGCAGAGAGAGAGAGAGAGAA	1541
Db	1333	GGAGAGAGCACTGGAAAAAGCAGCGGGAGCTTGAACGGCCAGAGAGAGAGAGAGAGAA	1332
Qy	1542	GGAGATCCAGAGAGCGCCAGAGCGCCCAAAACCGGAACTGGAAAGGCAGCGCAACTTGAA	1601
Db	1393	AGAAATTAGAGAGCGAGAGGCTGCAAAACGGGAACCTTGAAAGGCACGCAACTTGAGTG	1452
Qy	1602	GGAAACGAAACCGGAGACAGGAACCTCTGAATCAGAGAAACAGAGACAGAGAGGACACCGT	1661
Db	1453	GGAAAGGAATGAAAGGCAAGAACTACTAAATCAAAAGAAACAAAGAACAAAGAGACATAGT	1512
Qy	1662	GATCCTGAAGCAAGAGAGAAAGACTCTGAGAGTTGAGTTAGAAAGCTCTGAATGACAAAA	1721
Db	1513	TGTACTGAAGCAAAAGAAAAAGACTTTGGAAATTTGAAATTTGAAAGCTCTAAATGATAAAA	1572
Qy	1722	GCATCAGCTAGAGGAAAACTTCAGATATCAGGTGTCACTGGCAACCCAGAGGCAAGA	1781
Db	1573	GCATCAACTAAGAGGAAACCTTCAGAGTATCAGATGTCAATGCAACCCAAAGGCAAGA	1632
Qy	1782	AATTGAGCGACGAAACAAGCTACAGAGCTAAGAAATTTGGTGAATACCCACTTACACA	1841
Db	1633	AATTGAGCGCAAAACAAATTTAGAGGTTGAGAAATTTGCCGAATATACCCATCTACAGA	1632
Qy	1842	GCAGTTGACAGAAATCTCAGCAAAATGCTTGGAGAATTATTCAGAGAAACAGATACCTAG	1901
Db	1693	ACAATTAAGAGAACTCTCAGCAAAATGCTTGGAGACTTATTCAGAAAAACAGATACTCAA	1752
Qy	1902	TGACCACTTAAAAACAAGTCCAGAGCAACACTTTGCATAGAGACTCGCTTCTTAACCTCAA	1961
Db	1753	TGACCAATTTAAAAACAAGTTCAGAGCAACAGTTTGCACAGAGATTCACTTGTACCTTAA	1812
Qy	1962	AAGAGCTTTGGAAACCAAAAGAGCTGGCGCCGCGCAGACGCTCCGGAGACAGCTTGAAGAGT	2021
Db	1813	AAGAGCTTTAAGCAAAAGAACTAGCTCGGCGACACTTACGAGAACCAATCGATGAAGT	1872
Qy	2022	GGAGAGAGAGACAGAGCTCAAAAGCTTCAGAGAGATTGATGTTTCAACAAACCACTGAAGGA	2081
Db	1873	GGAGAAAAAAATCTAGATCTAAACTACAGAGAGATTGATATTTCAATTAACACTCAAGGA	1932
Qy	2082	ACTGAGAGAGATACATGACAAACAGCAACTCCGAAGCAGAGGTCCTCGAGGCGAGCGG	2141
Db	1933	ACTTAAGAGAAATACCAATTAAGCAACACTCCAGAACCAAAAGTCATGGAAGCTGAAACG	1992
Qy	2142	ACTGAAGCAGAAAGACAGAGAGAGAGAGACTCTGAGTTAGAGAACCAAAAGAAAGCGC	2201
Db	1993	ACTGAAACAGAAAGACACAGAACCAAAAGATCATAGAAATTTGAAAAACAAAAAGAAAGAGC	2052
Qy	2202	TCAAGAGAGACTTTCAGGAABAGGCAAGCAACTGTCTGAGAGCAATGTGCACAGACAGAG--GA	2258
Db	2053	CCAAAGACGAGCTCAGGAAGAGGCAAGCAAGCTGTGAGACATGTGCACAGAGAGAGCA	2112
Qy	2259	GCAGCAGCGCCCGCGAAACCCCAAGAGAGAGACAGACTGTAAGAGGAGAAAGAGTGTGAG	2318
Db	2113	GCATCAGAGACCAAGAAAACTCCACGAAGAGAAAAACTGMAAAAGGAGAGAGAGTGTCA	2172
Qy	2319	GAGAGAGAGCGCGGAGAGAGAGCCAAAGCCGGAATTCAAAGACAGACAGAGTGGCTTTT	2378
Db	2173	AAAGAAAGATGCGAGGAAAAAGGCAAAACAGAAAGCACAAGACAACTGGGTGGCTTTT	2232
Qy	2379	CCATCCGATCAGGAGCCAGCTTAAGCTGTGGCAACCAAGGCAACCTGGTCTACACAGAGAA	2438
Db	2233	CCATACACACCAAGAACCGAGTTAAAGCCAGCTGTCCAGGACACCTTGGTCCACTGCAGAAAA	2292
Qy	2439	AGGCGCGTTACCAATTTCTGCACAGAGAGTGTAAAGTGGTATATTAATCCAGGCGCTGTA	2498

Db	2293	AGGTCACTTACCATTTCTGACAGGAAATGTAAAGTGGTATTAACGGGCACTGTA	2352
Oy	2489	CCCTTTGAATCCAGAAAGTACGATGAGATACACATCCAGCCAGAGATATATGTCAT---	2555
Db	2353	CCCTTTGAATCCAGAAAGCCATATATGAAATCACTATCCAGCCAGAGACATATGATAGT	2412
Oy	2556	-----GGTGGATTAAAGCCAGACTGGAGAGCCAGGATGCTTGGAGAGACT	2603
Db	2413	TAAAGGGGAATGGGTGATATAAGCCAACTGGAGAACCCGGCTGGCTGGAGAGATT	2472
Oy	2604	GAAGGGAACAGCGGATGTTCCCTCGAAACTATGCAAGAAAAGATTCCAGAAATGAGCT	2663
Db	2473	AAAAGGAAAGCAGGGTGGTTCCCTCGAAACTATGCAAGAAAAGTCCAGAAAATGAGGT	2532
Oy	2664	TCCACATCCAGCCAAACCAATGACCGATGACATGTGCCCCCTGCCCCCAAACTGGCTCT	2723
Db	2553	TCCGCTCCAGTGAACCAATGACTGATTTCAACATGTGCCCTTGCCCCCAACTGGCTCT	2592
Oy	2724	GCGTGAAGACCCCTGCTCTTTGCACTGACCTTTCTGAGCCCTTCAGAACCCCCCAAA	2783
Db	2593	GCGTGAAGACCCCGCCCCCTTTGGCAGTAACCTTTGAGGCCCTTCAGACCCCTTAATA	2652
Oy	2784	CTGGGCAGACTTCAGTTTCCACGTGGGCCAGACGCTCAACAGAAAGCCAGAAACGAGAA	2843
Db	2653	CTGGCCGCACTTCAGCTTCCAGGGGCCACAGCAAGAAATGAGAAACAGAAACGAGATTA	2712
Oy	2844	CTGGGATACGTTGGGGCGGCTCAGGCTTCTCTGACCGTACCTAGTGTGGCCAGTTACGCA	2903
Db	2713	CTGGGATGCAATGGGCACGCCACGCTCTCTGACCGTTCCAAATGGCCGGCAGTTAAAGCA	2772
Oy	2904	GAGATCAAGCTTTTACCCAGCCACAGCACTGCTCTTCCCATTTCCCGTCTTGGGCA	2963
Db	2773	GAGGTCCGCTTTTACTCCAGCCACCGCCACTGCTCTTCCCGTCTCTGTGCTAGGGCA	2832
Oy	2964	GGGTGAAAAAGGTGGAAGGCTTCAACGCCAAGCCGTATTCCTGAGAGCCAAAAAGA	3023
Db	2833	GGGTGAAAAAGGTGAGGGGCTTCAAGCTCAAGCCCTATATCTTGGAGAGCCAAAAAGA	2892
Oy	3024	CAACCACTTAATTTTAACAAAAGTACGTCACTACCGTTCGTGAACAGCAAGACATGTG	3083
Db	2893	CAACCACTTAATTTTAACAAAATATATGTCATCACCGTCTGTGAACAGCAAGACATGTG	2952
Oy	3084	GTTGTTTGGAGAAATTCAGGTACAGAAAGGTTGTTCCCAAGTCTTACGTGAATCTAT	3143
Db	2953	GTTGTTTGGAGAAATTCAGGTACAGAAAGGTTGTTCCCAAGTCTTACGTGAATCTAT	3012
Oy	3144	TTTCAGGGCCGTAAAGAAATCCACAAGCACTGATCTGGCCCTTACTGAAGTCTGTAG	3203
Db	3013	TTTCAGGGCCCATTAAGAAAGTCTACMAAGCATGATTTGTGTTCTTTCAGAAAGTCTGTAG	3072
Oy	3204	TCTTAAAGAGATGCTTCCCCCGCCGCAAGCCAGCACTTCCCGAGAAAGATTTATTCG	3263
Db	3073	TCTTAAAGAGATGCTTCTTCACACACCAGAGCCGGTGTGTTGGAGAAAGATTTATTCG	3132
Oy	3264	CATGTAACAATACAGAGTCTGTAGCAAGAGATTTTAACTTTACAGCAAGGGAGTGTAT	3323
Db	3133	CATGTAACAATACAGAGTCTGTAGCAAGAGATTTTAACTTTACAGCAAGGGAGTGTAT	3192
Oy	3324	TGTGTTTACCAAGAAAGATGTGACTGTGTGAACGGGAACGTTGGCGACAACTCCGAGT	3383
Db	3193	TTTGTGTTTACCAAGAAAGATGTGACTGTGTGAACGGGAACGTTGGCGCGAGT	3252
Oy	3384	CTTCCCTTTCAATCAATGTGAGGCTTTAAAGATTCAAGGGCTCTGGAACTGCTGGGAAAC	3443
Db	3253	CTTCCCTTTCAATCAATGTGAGGCTTTAAAGATTCAAGGGCTCTGGAACTGCTGGGAAAC	3312
Oy	3444	AGGGAATTTAGAAAAAAACCTGAAATTTGCCAGGTTATTTGCTTCTTACGCTGTAATGG	3503
Db	3313	AGGGAATTTAGAAAAAAACCTGAAATTTGCCAGGTTATTTGCTTCAATACCGCCACGG	3372
Oy	3504	TCCGGAACAATCACTCCTGGCTCTTGGCAGCTGATTTTGATTCGGAAGAAAGAACCCAG	3563







Db 631 CCTCTTGTCTCTGTCTC-----CCACAGTTCACCCCTTGTCTTAACGGGGCTCTTGCA 684  
Qy 751 GTATPACAGCTCTGCGCTGCTTTGCGATCTGACGCCCATGGCCAAAGATTTCTCC 810  
Db 685 GTATATACAGTCTATCTGCTTTCCTGCTCTCT--GCCATCTGGCCAAAGATTTCTCA 741  
Qy 811 TTGACGAGATCTGGTCCAGGGTCAACAATTAAACAATTAAAGGACCAATCATTC 870  
Db 742 TTGCGCGCTCAGTTGCGGGTCCAGATTAACATACTTAAGGAGTCTGCTCTTC 801  
Qy 871 GATGTCGCCCGCCCTCCAGACGAGATGGGCTGTGCTCAGTCACTCAAGCTGAAA 930  
Db 802 GATGTCACGCCCTCCACTTGTGTTGAATGGCTGTGCTCAGTCTTCAATTTGAAG 861  
Qy 931 TACAGGAGTATTTCAACAGCCACGACAAATATAGTGGACATTAACAGTCCCGCAG 990  
Db 862 TACAGACATTTGTTCAACAGCCAGGACAAACCATGATGGAAATTAACAGTCTCCAG 921  
Qy 991 GCAAGAACTATTTCTCATGCAATCAAGTTTACCCAGGCTCAGCTGGCTTCAATATGAA 1050  
Db 922 GCAAGAACTATTTCTCATGCAATCAAGTTTACCCAGGCTCAGCTGGCTTCAATATGAA 981  
Qy 1051 CTTTCTGACATTTGATCAAGATGGAATACTCATGACAGAAATTTTACTAGCTATGCA 1110  
Db 982 CTTTCAACATTTGACCAAGATGGAATACTCAACGCGAATTTATATTTAGCTATGCA 1041  
Qy 1111 CTAAATGATGTTGGCATGTCTGTGCTGACGCACTGCGCGCTCTGCTCTCAATATCAT 1170  
Db 1042 TTATATAGTGTGGCATGTCTGCGCACGCACTTCTCTCAATCTGCTCTCAAGATATAT 1101  
Qy 1171 CTTCTCTCTCTGACAGAAAGTTGCTGCGGAGGAGTGTCCGTCATTAAGCTCTTCTCT 1230  
Db 1102 CCCCCTATCTTTTGAAGAGTTTCATCAGGAGTGGTTATTCATTAATGACTCTGTGTCT 1161  
Qy 1231 GTGATCAGAGGCTGCTGAGAGCCGCTGTCAGAGGATGACGACCAAGAGAAA 1290  
Db 1162 GTTATCAGCGCTTGGCGGAGAACCAAGAAAGAGAGGACCAAAACGCGGACAAAG 1221  
Qy 1291 CTGCTGTGACATTTGAAGATTAAGACGGGAGAACTTTCAGCGAGGCTATGAGCTG 1350  
Db 1222 CTCCCAATGATTTGAAGACAAAGAAAGGAGAACTTTGAGCGAGAACTTAAGTTA 1281  
Qy 1351 GAGAAAGCGCGCCAAAGCGCTTGTGAGCAGACGCGCAAAAGAGCGGTTGGCTCAG 1410  
Db 1282 GAAAGAGAGAGACAGGACCTTTTACAGACAGCGCAAAAGAGAGAGAGTTGGCTCAG 1341  
Qy 1411 CTGAGAGCGCGCGAGCAGAGAGAGAAAGAGCGGAGCGCCAGAGAGCGGAGCGG 1470  
Db 1342 CTGAAAGGGGCAAGAGAGAGAAAGAAAGAGAGCGGCAAGATCAAGAGCGGAGCGG 1401  
Qy 1471 CAGCTGAGCTGAGAGACAGCTGGAAGAGCGGAGCTGAGCGGCGAGCGAGAGAG 1530  
Db 1402 CAGGAAAGCTTGAAAAAGAGCTGGAAGAAAGAGAACTGAGAGCGGCAACGGAGAG 1461  
Qy 1531 GAGAGAGAGAGAGATGAGAGCGCGGAGCGGCAAAAGCGGAGCTGGAAGCGAGCGA 1590  
Db 1462 GAGGCGAGAAAGAAATCGAAAGAGAGCGGCAAAAGAGAACTGGAAGAGCGAGCGG 1521  
Qy 1591 CAACCTGAATGGAACGGAACCGGAGCAGGAACCTCCGAACTCAGAGAAACAAGAGAGAG 1650  
Db 1522 CAGTTGAGTGGAAACGAAACAGAAAGCGAGAACTCTTAATCAAGAAATAGAGAGAA 1581  
Qy 1651 GAGGAGACCGTGTCTGGAAGAGCAAGAGAGAGAGCTGAGAGTTGAGAGTTGAGAGCT 1710  
Db 1582 GAGGACATTTGTTCTTAAAAAGCAAGAAAGAAAGACTTGAAGTTGAGCTGGAAGCTCTA 1641  
Qy 1711 AATGACAAAAAGCATCAGCTTGAAGAGAAACTTGAAGATATCAGGTTCAGCTGGAGCC 1770  
Db 1642 AATGATAAAAAGCAACCAATTTGAAGAAAGCACTCAAGATATCCGATGCTCACTACT 1701  
Qy 1771 CAGAGGCAAGAAATTTGAAGAGCAAGAACTGAGAGAGCTTAAGAAATTTGCTGAATCACC 1830  
Db 1702 CAACGGGCGAAGAAATGAAGAGCACTAACAAATCCAGGGAACTGAGGATCGAGAGATCA 1761

Qy 1831 CACTTACAGCAGAGCTTGAGGAATCTCAGCAATGCTTGGAAAGCTTATTTCCAGAGAA 1890  
Db 1762 CACTTGCAGCAGAGCTTTCAGGAATCTCAGAGCTGCTTGGAAAAATGATCTCTGAGAAA 1821  
Qy 1891 CAGATCTCATGACAGTTTAAAAAAGCTCAGAGAAAGCTTTGCTATAGAGCTGCTT 1950  
Db 1822 CAGTCCCTTATGATCAATTAAGCAGGTTTCACAAACAGTTTTCACAGAGATTCCTTG 1881  
Qy 1951 CTTACCTCAAAAGAGCTTGGAAAGCAAGAGAGCTGGCCCGGAGAGAGCTCCGAGAGAG 2010  
Db 1882 CTTACCTCAAAAGAGCTTGGAAAGCAAGAGAGATTTGAGCGCAGAGCTCCAGATCAG 1941  
Qy 2011 CTGAGCAGCTGAGAGAGAGACAGAGTCAAGCTGAGAGAGATTTGATTTTCAACAC 2070  
Db 1942 TTGATGAGAGTGAAGAAAGAAAGCGAGCTTAAGCTTGAAGAAATCGAGTATTAATAC 2001  
Qy 2071 CAGCTGAAGAACTGAGAGAGATACATAGCAAAACAGCAACTCCAGAAAGAGAGTCCCTG 2130  
Db 2002 CAATTTAAAGAACTGAGAGAAATTTGACACAGCAGAGTTCCAGAAACAGAGAGACTTT 2061  
Qy 2131 GAGCAGCGGCACTGAAAGCAAGAAAGAGAGAGAGAGCTTGAGTTAGAGAGCA 2190  
Db 2062 GAAACTGAGAAAGATCAAGCAGAAAGAGCTGGAACGAAAGAAACAGCGAGCTGATTAAGCTG 2121  
Qy 2191 AAGAGAGCGCTCAGAGAGCGAGTTCAAGAAAGGAGCAAGCAATGGCTGAGCATGTGCA 2250  
Db 2122 AAGAGAGAGCAAAAGAGCGAGATGCTAGAGAGAGCAAGCTTTGGCAGAGATCCGCTGA 2181  
Qy 2251 CAGAGAGAGCAGCAGCGCCCGGAGAAACCCAGAGAGAGCAGACTGAAGAGGAGAAAG 2310  
Db 2182 CAGAGAGAGAGAACTTAC-----AAATTCAGAGATCAGGAGAAAGAAAGGAGAA 2235  
Qy 2311 AGTGCAGAGAAAGAGAGAGCGGAGAGAGAGCAAGCGGAAATGCAAGCAAGAT 2370  
Db 2236 TCCGTTCAAGAGTGCAGAGTGAAGAG-----AAACCAAGATCCAGAGAAACCAAC 2289  
Qy 2371 CGGCTTTTCAATCCGATCAGAGAGCAGCTAAGCTGACCAAGAGAGAGAGAGAGAGCTTAC 2430  
Db 2290 AAACCGTTCCAGCAACCTCCAGAGCTGAGAAATCAAGAGGCGAGATTCCTCGATGAAT 2349  
Qy 2431 ACAGAGAAAGCGCGCTTACATTTCTGCAAGAGAGCTGAAGAGTGAATATTAACGA 2490  
Db 2350 ACAGAGAAAGCGCGCTTACATTTCTGCAAGAGAGAGAGAGAGTGAATATTAACGA 2406  
Qy 2491 GCGCTGTACCCCTTGAATCCAGAAAGTCAAGATGAGATCACTATCCAGCGAGAGATATA 2550  
Db 2407 GCATCTACCTTTGATGCGGAGAGTATGATGAATCACTATTTGAGCTGAGGACATTT 2466  
Qy 2551 GTCATGTGATGAAAGCCAGACTGAGAGAGCGCAGATGAGCTTTGAGAGAGAGCTGAAGGG 2610  
Db 2467 ATCATGTGATGAAAGCCAG 2526  
Qy 2611 AAG 2670  
Db 2527 AAAAGCGGCTGTTTCTGCGCACTATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2586  
Qy 2671 CCAGCGAAACAG 2730  
Db 2587 ACAACTAAACAG 2638  
Qy 2731 ACCCTGCTCTTTGCGAG 2790  
Db 2639 -CTTCCCTGTTGAG 2697  
Qy 2791 GACTTCAGTTCCAGCTGAG 2850  
Db 2698 GACTTCAGCTGAG 2757  
Qy 2851 ACATGGGCGAGCTCAGCTCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2910  
Db 2758 ACATGGGCGAGCTCAGCTCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2817



QY 426 TGNCTTAGCAAAATATAGGGGCTGAGGAGCATGAATTAAGATGAAGATGATCACT 485  
 Db 514 TGTATTAAGACAAATATGGGCACTAGCTAGATGAATTAATGATGAAGATGATCACT 573  
 QY 486 GGAATTTTCATGACATGAAGCTTATCAAACTGAAGCTTACAAAGATATCACTCCCTC 545  
 Db 574 GGAATTTTCATGATGAAGCTTATCAAACTGAAGCTTACAAAGATATCACTCCCTC 633  
 QY 546 CAAATTTCCCTGCTATGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 605  
 Db 634 TGCACTTCCCTGCTATGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 691  
 QY 606 TATAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 665  
 Db 692 ----- 691  
 QY 666 CATTCAGTGTGGAATGTCTCCACCCTTATCTTCTCTCTCTCTCTCTCTCTCTCTCT 725  
 Db 692 ----- GCAAGCTGTGCT 702  
 QY 726 TCCCTTGGCTAAGGGGCTCTCCCTCATACAGCCCTGCTGCTGCTGCTGCTGCTGCTG 785  
 Db 703 CCCCCTGGCTAAGGGGCTCTCCCTCATACAGCCCTGCTGCTGCTGCTGCTGCTGCTGCT 762  
 QY 786 AGGCACATGGCCAAAGATCTTCTCTTCAAGATCTGCTCAAGGCTCAAAATTAACAC 845  
 Db 763 AGGCACATGGCCAAAGATCTTCTCTTCAAGATCTGCTCAAGGCTCAAAATTAACAC 822  
 QY 846 TAAGTTACAGAAAGCAATCATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 905  
 Db 823 TAAATTAAGAAAGCAATCATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 882  
 QY 906 TGTGCTCAGTCAATCAAGGCTGAATTAACAGGCTTATTAACAGGCTCAAAATTAAC 965  
 Db 883 TGTGCTCAGTCAATCAAGGCTGAATTAACAGGCTTATTAACAGGCTCAAAATTAAC 942  
 QY 966 GAGTGAACATTAACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1025  
 Db 943 GAGTGAACATTAACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1002  
 QY 1026 GGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1085  
 Db 1003 GGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1062  
 QY 1086 AGAAGATTTATCTAGCTATGCACTTAATGATGTTGCCATGCTGCTGCTGCTGCTG 1145  
 Db 1063 AGAAGATTTATCTAGCTATGCACTTAATGATGTTGCCATGCTGCTGCTGCTGCTG 1122  
 QY 1146 GCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1205  
 Db 1123 ACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1182  
 QY 1206 GATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1265  
 Db 1183 TATATCTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1242  
 QY 1266 GATGAGACAGCAGC---CAGAGAGAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1322  
 Db 1243 AGATGAACAACAATTAAGAAAGAAATTAACCTGTAACCTTGAAGATTAAGAAAG 1302  
 QY 1323 GAATCTGACAGGAGCAGTGTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1382  
 Db 1303 GAACTTTGAAGCTGAGCACTGGAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1362  
 QY 1383 GCGCAAG 1442  
 Db 1363 GCGCAAG 1422  
 QY 1443 GGAAGCTGAG 1502  
 Db 1423 TGAGCGGCAAG 1482  
 QY 1503 GCGGAGAGCTGAG 1562

Db 1483 GCGGAGAGCTAGAACAGCAG 1542  
 QY 1563 CGCAAAACGGGAACTGGAAGAGCAGAGCAAACTTGAATGAGGAACGGAACCGGAGACAGCA 1622  
 Db 1543 TGCAAAACGGGAACTTGAAGAGCAGAGCAAACTTGAATGAGGAACGGAATCGGAGAGAG 1602  
 QY 1623 ACTCTGGAATCAG 1682  
 Db 1603 ACTCTGGAATCAG 1662  
 QY 1663 GACTGTGAGTTGAGTTAGTAAGCTTGAATGACAAAGACATCAGTGAAGAGAGAGAGAG 1742  
 Db 1663 GACTTTGGAATTTGAATTTAGAGAGCTTAATGATTAAGAAACATCACTAGAGAGAGAG 1722  
 QY 1743 TCAGATATCAGAGTGTGCACTGCGAACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1802  
 Db 1723 TCAAGATATCAGAGTGTGCACTGCGAACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1782  
 QY 1803 TAGAGAGCTAAGAAATGCTGGAATCACCACCTTAACAGAGAGAGAGAGAGAGAGAGAG 1862  
 Db 1783 TAGAGAGTGAAGATGCTGGAATCACCACCTTAACAGAGAGAGAGAGAGAGAGAGAG 1842  
 QY 1863 AATGCTTGAAGACTTATTCAG 1922  
 Db 1843 AATGCTTGAAGACTTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1902  
 QY 1923 GCAGAAAGCTTGAAG 1982  
 Db 1903 GCAGAAAGCTTGAAG 1962  
 QY 1983 GCTGCGCCGAG 2042  
 Db 1963 ACTAGCTGCGAG 2022  
 QY 2043 GCTGCGAG 2102  
 Db 2023 ACTAG 2082  
 QY 2103 ACAGCAACTCCAG 2162  
 Db 2083 GCAACAATCCAG 2142  
 RESULT 11  
 HSU61166 3241 bp mRNA linear PRI 23-JUL-1996  
 LOCUS  
 DEFINITION Human SH3 domain-containing protein SH3P17 mRNA, complete cds.  
 ACCESSION U61166  
 VERSION U61166.1 GI:1438932  
 KEYWORDS  
 SOURCE  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 3241)  
 Sparks, A.B., Hoffman, N.G., McConnell, S.J., Fowlkes, D.M. and  
 Kay, B.K.  
 Cloning of ligand targets: systematic isolation of SH3  
 domain-containing proteins  
 Nat. Biotechnol. 14 (6), 741-744 (1996)  
 JOURNAL MEDLINE  
 PUBMED 9630982  
 REFERENCE 2 (bases 1 to 3241)  
 Pirozzi, G., McConnell, S.J., Uveges, A. and Fowlkes, D.M.  
 Direct Submision  
 Submitted (18-JUN-1996) CYTOGEN Corp., 307 College Road East,  
 Princeton, NJ 08540, USA  
 FEATURES  
 source  
 1..3241  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /tissue\_type="bone marrow"

Query Match	Best Local Similarity	Score	DB	Length
24.6%; 69.8%;	24.6%; 69.8%;	1252.6;	DB 9;	3241;
Matches 2132; Conservative 0; Mismatches 604; Indels 317; Gaps 20;	Matches 2132; Conservative 0; Mismatches 604; Indels 317; Gaps 20;			
Query 2112	CCAGNAGCAGAGGTCCTCTGAGGCGACGCGACTGAAAGCAAGAAAGCAGAGAGAGAG	2171		
Db 21	CCAGAGCAAAAGTCATGAGAGCTGAAAGCACTGAAACAGAAAGAAACAAAGACGAAAGT	80		
Query 2172	CCTGAGATTAGAAAGCAAAAGAGAGACGCTCAGAGACGAGTTCAAGAAAGGACACAGCA	2231		
Db 81	CATAGAAATTAGAAAAACAAAAAGAAAGCCCAAGACGAGCTCAGAAAGGACACAGCA	140		
Query 2232	ATGGCTGAGACATGTCGACGACGAG---GAGGAGCCAGCCCGGAAACCCACAGAGCA	2288		
Db 141	GTGGCTGAGACATGTCGACGAGGAGGACGAGATCAGAAACCAAGAAACCTCAGAGAGA	200		
Query 2289	GCACAGACTGAAGAGGAGAAAGCAGTGTGCAAGAAAGAGGCGGAGAGAGAGCCAAAGCC	2348		
Db 201	GGAAAACTGAAAAAGGAGAGAGAGTGTCAAAAAGAGAGATGGCGAGGAAAAAGGCAACA	260		
Query 2349	GGAAATGCAAGCAAGCAGAGAGTGGCTTTTCATCCGATCAGAGCCAGCTAAGTGTGC	2408		
Db 261	GGAAACACAGACAGACAGTGGTGGCTTTTCATCAACACCAAGAAACCGCTAAGCCAGC	320		
Query 2409	CACCCAGGACCCCTGTGTCTACACACAGAGAAAGCCCGCTTACCATTTCTGACAGAGAG	2468		
Db 321	TGTCCAGGACCCCTGTGTCTACACACAGAGAAAGGTCACCTTACCATTTCTGACAGAGAAA	380		
Query 2469	TGTAAAGTGTATTTATCCGAGCGGTATACCCCTTTGATTCAGAAAGTCAAGATAGAT	2528		
Db 381	TGTAAAGTGTATTTATCCGAGCGGTATTTACCCCTTTGATTCAGAAAGTCAAGATAGAT	440		
Query 2529	CACCATCCAGCAGAGAGATATAGTATGATGGATGAAGCCAGACTGAGAGCCAGAGATG	2588		
Db 441	CACCATCCAGCAGAGAGATATAGTATGATGGATGAAGCCAAACTGGAGAACCCGGCTG	500		
Query 2589	GCTTGGAGAGAGCTGAAAGGAGACGGGATGTTCCCTGCAAACTATGACAGAAAAAGT	2648		
Db 501	GCTTGGAGAGAGATTTAAAGGAGAGACAGGGTGTTCCTCTGCAAACTATGACAGAAAAAT	560		
Query 2649	TCCAGAAAAATGAGTTCCCACTCCAGCCCAACACAGAGCCGATCTGACATCTGCCCTCG	2708		
Db 561	CCAGAAAAATGAGTTCCCGCTCCAGTGAACACAGATCTGATTTCAACATCTGCCCTCG	620		
Query 2709	CCCCAAACTGAGCTGCGTGAAGACCCCTGCTCTTTGACAGTGAACCTCTTCTGAGCCCTC	2768		
Db 621	CCCCAAACTGAGCTTGGCGTGAAGACCCCGCCCTTTGGGAGTAACTCTTTCAGAGCCCTC	680		
Query 2769	CACAAACCCCAACACTGGGACAGACTTCAAGTTCAAGTGGCCACGAGCTCAACAGAGAA	2828		
Db 681	CACGACCCCTTAATTAAGTGGGCGACTTCAAGCTTCAAGTGGCCACACGACGCAATAGAA	740		
Query 2829	GCCAGAAAGGACAACTGGGATACGTGGGCGGCTCAGAGCTTCTGTAACCTTAAGTGC	2888		
Db 741	ACCAGAAAGGATTAAGTGGGATACGTGGGCGGCGCCAGCCCTCTCTTCAACGTTCAAGTGC	800		

Qy	2889	TTGGCAGTTTACGGCACAAGATCAGCCCTTACCCACGACACAGCACTGGCTCTCCCATC	2948
Db	801	CGGCGCATTTAAGCGAAGGTCGCGCTTTACTCAGCAGAGCCACTGGCTCTCCCGCTC	860
Qy	2949	TCCCGTCTGGGCAGAGGTGAAAAGGTGAGAGGGCTACAGCGCAAGCCCTGTATCCCTG	3008
Db	861	TCCTGTCTAGGCGCAGGTGAAAAGGTGAGGGGGCTACAGCTCAAGCCCTATATCTTG	920
Qy	3009	GAGAGCCAAAAAGACACACCTTAATTTTAAACAAAGTACGTCATCACCCTTCTGGA	3068
Db	921	GAGAGCCAAAAAGACACACCTTAATTTTAAACAAAAATGATGTCATCACCGTCTGGA	980
Qy	3069	ACACAGACATGTGGGTGGTTTGAGACAGTTCAGAGTCGAAAGGTTGTTCCCAATC	3128
Db	981	ACACAGACATGTGGGTGGTTTGAGAGATTTCAAGTTCAGAGTCAAGAGGTTGGTCCCAATC	1040
Qy	3129	TTACGTGAAACTGATTTCAAGGCCCGTAAAGAAATCCACAGACATCGATCTGCGCTAC	3188
Db	1041	TTACGTGAAACTGATTTCAAGGCCCGTAAAGAAAGTCTACAGACATCGATCTGCGCTTC	1100
Qy	3189	TGAAAGTCTGCTAGTCTAAAGAGATGCTTCCCGCGCCAGACCATTCGCCG	3248
Db	1101	AGAAAGTCTGCTAGTCTAAAGAGAGATGACTCTCCAGAGCCAAAGCGGTGTTCGGG	1160
Qy	3249	AGAAGATTATTGCACTGACACATACAGAGATTCTGAGCAAGAGATTTAACTTTCA	3308
Db	1161	AGAA-----	1164
Qy	3309	GCAAGGGATGTGATTTGTGTGTTACAAAGAAAGTGTGACTGTGTGACGGGACGGTGG	3368
Db	1165	-----	1164
Qy	3369	CGACAAGTCCGAGTCTTCCCTCTTAAGTATGAGGCTTAAAGATTCAAGGGCTCGG	3428
Db	1165	-----	1164
Qy	3429	AACTGCTGGAAAAACAGGGAGTTTAGGAAAAAAACCTGAAATTGCCAGTTATTGCTTC	3488
Db	1165	-----GAAATTGCCAGTTATTGCTTC	1188
Qy	3489	CTACGCTGTCTAGTGTCCGACACACTCACCTGTGCTCTGGGCACTGTATCTGATCCG	3548
Db	1188	ATACACCGCCACCGCGCCCGACGAGCTCACTCTCGCCCTGTGCACTGATTTTATCCG	1247
Qy	3549	GAAAAAGAACCCAGGTGATGTGTGGGAGAGAACTGCAAGCTCGAGGGAAAAAGCGCA	3608
Db	1248	AAAAAGAACCCAGGTGATGTGTGGGAGAGAGAGCTGCAAGCACTGTGGAAAAAGCGCA	1307
Qy	3609	GATAGGGTGTTCACAGCAATTAATGTCAAACTTTAGCCCCCGGAACAAGCAAAATCAC	3668
Db	1308	GATAGGGTGTTCACACTTAATTAATGTAAAGCTTTAAGCCCTGGGACAGCAAAATCAC	1367
Qy	3669	CCCAACTGAGCTTACCCAAGACCGCAGTGCACGACGACGAGTGTGCCAGTGTATCGGATGTGA	3728
Db	1368	TCCAAACGAGCCACTTAAGTCAACAGGATTTAGGGAGTGTGCAGGTGATTTGGATGTGA	1427
Qy	3729	CGATTACACCGCCAGAAACGATGACGAACCTAGCCTTCAGCAAAAAGCCAGATCATCAACGT	3788
Db	1428	CGATTACACCGCGAGAAATGATGATGAGTGTGCTTCACACMAAGGCCAATCATCAACGT	1487
Qy	3789	CCTCAACAAGAGACCCCGACTGTGTGAAAAGAGAAATCACTAGTGGCAAGTTGGGCTCTT	3848
Db	1488	CCTCAACAAGAGACCCCTGACTGTGTGAAAAGAGAAATCAATGCAAGTGGGGCTCTT	1547
Qy	3849	CCCATCCAAATTAATGAAGCTGACACACAGCATGAGACCCAGCAGCAATGATCATATNG	3908
Db	1548	CCCATCCAAATTAATGAAGCTGACACACAGCATGAGACCCAAAGCAGCAATGATCATATNG	1607
Qy	3909	TTGTCCATCCCCCTCAGGCTTGAAGTCTCTAAAGAGACCACTATCCATATCACTG	3968
Db	1608	TTGTCCATCCCCCTCAGGCTTGAAGTCTCTAAAGAGACCACTATCCATATCACTG	1667

Oy	3969	CCAAAGGAGATGATGGGAGATGACAGCCTTATGATCATGATGCTTACACATGATCACCTACT	4028
Db	1668	CCCAAGAGGAGATGATGGGAGATGACAGCCTTATGATCATGATGCTTACACATGATCACCTACT	1722
Oy	4029	GCTTCTGAG-TAGAAAGACTCACTGACAGACGATTACTCTCATTTGACCTTAGTTGCAT	4087
Db	1728	GCTTCTGAGTTAGAGAACTCACTGCAAAACGTTTACCCCATTTAACCTTAGTTGCAT	1787
Oy	4088	GTGATCGAAATAGTCTGAGTCACTGCGGCGAGAGGCGAAGACAA-----TTGAGAACTGC	4143
Db	1788	GTTAACCCCAAGTTTGAATTAACTCTGGCAAAAATGAAACCAAAATTTTCATPAAACCC	1847
Oy	4144	ACAGGTTGGTGGGTCCTTTTGGGGCTTTCTCTAGTCACTGACTGA-----CCGG	4193
Db	1848	ACGGGGTAGTGGGTCCCTTTGTGTGGCTTTCCCTAATTAACTCATAATGAATTTCCCC	1907
Oy	4194	CCCCCTTCAACAGGGGCGCTTCAATAGTTTAAATTTATTTTAA-----TGT	4244
Db	1908	CACCTTGGCCACAGGGGCGCTTCAAAATATTTTAAATTTAAATTTAAAAAAATTTAT	1967
Oy	4245	GTAATTTAGCCTTTTAAATAAATATCTCAATCAATTACTCTTGGCTATTTGGTTT--	4302
Db	1968	TGATTTAACCTTTTAAATAAACCAAAATATAATTAATTACTCTTGTCTATTTGGGGTTGG	2027
Oy	4303	-ACAAAAACCCCACTATCAAGAGTGCCTGTCTGCGAGATTAATAATGCTGTTCCGGG	4361
Db	2028	CAAAAAGACCCCTATTCAAAAGGAAATGCGCTGTGCTATTAATAAAATATGTTCCAAA	2087
Oy	4362	CGTACCCTAACTGAGAGCTTGCCTGTACCT-----TTGCGCTTTGTCGAGTTCGCCAC	4416
Db	2088	ATTTTCCATTAACCGGTBAACCTGAATGTCTTCTTCAATTTTGTCTCCCGCTTTACCAAC	2147
Oy	4417	-----CACATGTGTATGTTTGGGCGTGT-----CC	4442
Db	2148	CTAAATTTGCTGCACACTTGTGGGGGCGCTTGGTATTTTATTTTTCCTCCCCCTTCAACC	2207
Oy	4443	CTGCGGTAGACACAGAGAAATGGGTGATCT-----GTTTCAAAATGCG	4489
Db	2208	CTTCAAAATTTCTCCCAAAAAGAAAGTTTCTCTCTCCCCCCCGCTATTAATTAACCA	2267
Oy	4490	TATGTAGACTGAGCTGACTATGAAAGGGGTATGTGTTGCTGTGATCAATCACTGATCC	4549
Db	2268	CACTTAAACTGATACCCCCCAACCCCCCTCTAAAAAAAACCAATGGTTTTTGTGT	2327
Oy	4550	TGTGCGCATGTACATCTGTATGCCAAGAAAGTAGTACTTCTCTCATAGCTAAACCCACA	4609
Db	2328	GACCACTTAAAGCATCTCTGTCTCAATACAAATTTCTTTTTTCCAGGGCAAAAGTATTA	2387
Oy	4610	CCGTATACAGTGCCTC--TCACTACTGCATTCATTTTACTTTTGC-ACAGTACCTGTGA	4665
Db	2388	CCTTGTATGATGCTCTAATCATATTTGGCATTTATTTTATTTTGGCAACGTATCACTTGTGA	2447
Oy	4666	GCCACTTGAGAGAGAC-CCATGTTTCCGTTTGTGCTCAGATGTACTAGTTGTCCCT	4724
Db	2448	GCCACATAGAGAAACACACTGTGTGTTTTTGTGCTGTCAAGATTTATCTGTGTGAATGGT	2507
Oy	4725	GTTTTGTTTTTATTTTTCATATCTGGCATGTCTTCAACACATAAAGTAAGTAACGCCACAC	4784
Db	2508	GTTTTGTTTTGGGTTTTTAAATTTTGTGCTGTTTGTGATGACATATAATCACTATGACAAAC-AC	2566
Oy	4785	TGCCAGCGGTTTACGATTCATCAAGTACCAACCGTCTC-----TTAGTCTGTGTACGTGAAG	4839
Db	2567	CACGTAGGTCGTTACGATTCACATCCAGTATCTGTCAGTCTCTGTTACTAGTAAG	2626
Oy	4840	TTTATTTCAAGTGTGCTTTTATGAA-----TATCTTGAACAAGTATCTTCTTGACAG	4893
Db	2627	TTTTTATTCAGTTACTTTTCAATGGAAGACCTATTTTGAACAAGTAAATTTCTTGACAG	2686
Oy	4894	AAAAGATATATAGAAGTCTCCCTGCATTAATTTTCCAGAGTTTACA--TTTTTAACT	4950
Db	2687	AAAAGATATATAGAAGTCTCCCTGCATTAATTTTCCCAAAAGTTAAAAATTTTTTAAATA	2746
Oy	4951	AGACTGTGGGGTTGCTACAGATTAAT-ATGAATATGGCGCTCTGCTGCTGTGTGTGT	5008

Db 2747 GGACTGCGGAATTTTAAAGATTATGAAATGAGCTCAGGCTCGTTGGGGT 2806  
Qy 5009 TAACCTGCTAGCTGAAGCCGCTGCTAGATATTAGTTGGAAGTCG 5061  
Db 2807 AAGAAAGCTGTAGGGAAAGCCCTTTGTTTAAACATAGTGAAG 2859

RESULT 12						
AF169621	AF169621	1133 bp	mRNA	linear	ROD 12-AUG-1996	
LOCUS	Mus musculus	Ese1 protein mRNA, partial cds.				
DEFINITION	AF169621					
ACCESION	AF169621.1	GI:5731280				
VERSION						
KEYWORDS	Mus musculus.					
SOURCE						

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus  
1 (bases 1 to 1133)

AUTHORS	TITLE
Isyda, L.O., Kvasna, S.M., Skripkina, I.I., Anoplienko, O.V.	Mouse homologues of human chromosome 21 genes
Slavov, D., Tassone, F., Rynditch, A.V. and Gardiner, K.	

REFERENCE  
2 (bases 1 to 1133)  
AUTHORS  
Tsyba, I.O., Kvasha, S.M., Skripkina, I.Y., Anoprienko, O.V.  
Slavov, D., Tassone, F., Rynditch, A.V. and Gardiner, K.

**JOURNAL** Submitted (15-JUL-1999) Department of Molecular Oncogenetics,  
Institute of Molecular Biology and Genetics of National Academy of  
Sciences of Ukraine, 150, Zabolotnogo str., Kiev 252627, Ukraine

FEATURES  
SOURCE

```

/organism="Mus musculus"
/dev_xref="taxon:10090"
/dev_strage="15-day embryo"
<1.749
CDS

```

```

/note="similar to Homo sapiens interectin"
/codon_start=3
/product="Esel protein"
/protein_id="AAB48848.1"
/db_xref="GI:5731281"
/translation="VRKSTIDRGPTESPASLRKVASPAKPAIDGEEFIAMTYESS
EGDITFOOGDIVIVTKRSGDMMWTGVCRSGVPSNRYRLKDSGSGTACKTSLG
KELIAQVASYATGAPQLFLAPGQILIRKNPGGWMGELOARGRKROIQMPANPY
VKLSPGVSKITPELTKTAVOBACGVIGMDVTAQNDELAFSKQIINVLNKEDP
DMWKEBSGVGGLFPNSYVTKLTIDMDPSQQ"

```

BASE COUNT	316 a	276 c	285 g	256 t
ORIGIN				
Query Match		21.8%	Score 1110;	DB 10; Length 1133;
Best Local Similarity		99.7%;	Pred. No. 3.1e-244;	
Matches 1133; Conservative	0;	Mismatches	0;	Indels 3; Gaps 2;

3152 CCGTAAGGAAATCCACAAGCATCGATTCTGGCCCTACTGAAAGTCTCTGTAAGCTTAAGA 3211

Db 1 CCGTAGGAAATCCACAAGCATCGATCTGGCCCTACTGAAAGTCTGTGTCGTCTAAGA 60

QY 3212 GAGTGGCTTCCCCGGCCGACCAAGCCATTCGCCGAGAAGATTATTGCGATGTACA 3271

Db 61 GAGTGCTTCCCGCGCGCCAAAGCCAGCCATTCCCGGAGAGAGTTATTGCCATGAC 120

QY 3272 CATACGAGAGTTCTGAGCAAGAGATTAACTTTACCTTTCAGCAAGGGATGTGATTGTGGTTA 333

Db 121 CATACGAGAGTTCTGACCAAGGAGATTAACTTTCAGCAAGGGATGTGATGTGCTTA 180

DY 3332 CCAGAAAGATGGTAGCTGGTGGAGGGAAGGTGGTGACCAAGTCCTGAGATCCTCCCTTT 3339

DH 181 CCGAATAAGCATGCCTGTAATTGCCTGAGAAGGCTGGAGGACAAGCTCCGAGACTGCTCCCTT 240

[illegible]

Db 241 CTAAGTATGTTAGGCTTAAGATTCAGAGGCTCTGGAATCGTGGGAAAACAGGAGTT 300

QY 3452 TAGAGAAAAAAGCTGAATTTGCCAGATTATGCTTCTACGCTGTCTACTGTGCTCCGAAC 3511  
 DB 301 TAGAGAAAAAAGCTGAATTTGCCAGATTATGCTTCTACGCTGTCTACTGTGCTCCGAAC 360  
 QY 3512 AACTCACCCTGCTCTCTGGGAGCTGATTTCTGATCCGAAAAAAGAACCCAGGTGATGGT 3571  
 DB 361 AACTCACCCTGCTCTCTGGGAGCTGATTTCTGATCCGAAAAAAGAACCCAGGTGATGGT 420  
 QY 3572 GGGAGAGAGAACTGCAAGCTCCGAGGAAAAAGGCGAGATAGGATGTTTCCAGCAAAAT 3631  
 DB 421 GGGAGAGAGAACTGCAAGCTCCGAGGAAAAAGGCGAGATAGGATGTTTCCAGCAAAAT 480  
 QY 3632 ATGTCAAACTTTCTAGCCCCCGGAACAAATCAACCCCACTAGCTACCCAAAGACCG 3691  
 DB 481 ATGTCAAACTTTCTAGCCCCCGGAACAAATCAACCCCACTAGCTACCCAAAGACCG 540  
 QY 3692 CAGTCAGAGCAGAGATGTCGAGGTATGGGATGATGACATTAACCCGCCAGAACGATG 3751  
 DB 541 CAGTCAGAGCAGAGATGTCGAGGTATGGGATGATGACATTAACCCGCCAGAACGATG 600  
 QY 3752 AGCACTAGGCTTCAGCAAAAGGCGAGATCATCAAGCTCTCAACAGAGAGACCCGACT 3811  
 DB 601 AGCACTAGGCTTCAGCAAAAGGCGAGATCATCAAGCTCTCAACAGAGAGACCCGACT 660  
 QY 3812 GGTGAGAGAGAGAGTCAAGTGGGCAAGTTGGGCTCTTCCATCCAAATTAATGAAGCTGA 3871  
 DB 661 GGTGAGAGAGAGAGTCAAGTGGGCAAGTTGGGCTCTTCCATCCAAATTAATGAAGCTGA 720  
 QY 3872 CCACAGACATGACCCCGAGCAGCAATGAATCATATGTTGTCATCCCCCTCAGGCTT 3931  
 DB 721 CCACAGACATGACCCCGAGCAGCAATGAATCATATGTTGTCATCCCCCTCAGGCTT 780  
 QY 3932 GAAAGTCTCAAGAGAGACCATATCCCATATCACTGCGCAGAGGATGATGGAGATGC 3991  
 DB 781 GAAAGTCTCAAGAGAGACCATATCCCATATCACTGCGCAGAGGATGATGGAGATGC 840  
 QY 3992 AGCCTTGATCATGATGATGTCGACATGATCACTACTGCTTCTGATGAAGAACTCAC 4051  
 DB 841 AGCCTTGATCATGATGATGTCGACATGATCACTACTGCTTCTGATGAAGAACTCAC 900  
 QY 4052 TGCAGAGCAGTTAATCTCATTTGATGATGATGATGATGATGATGATGATGATGATG 4111  
 DB 901 TGCAGAGCAGTTAATCTCATTTGATGATGATGATGATGATGATGATGATGATGATG 960  
 QY 4112 CGGCGAGAGCAGAGAAATGCAAGATGCAAGGATGATGATGATGATGATGATGATGATG 4171  
 DB 961 CGGCGAGAGCAGAGAAATGCAAGATGCAAGGATGATGATGATGATGATGATGATGATG 1020  
 QY 4172 CTTAGTCACTCAGACTGACCGGCGGCTTCAACGAGGCGCTTTCAATAGTTTAAAGAT 4231  
 DB 1021 CTTAGTCACTCAGACTGACCGGCGGCTTCAACGAGGCGCTTTCAATAGTTTAAAGAT 1079  
 QY 4232 TATTTTAAATGATATTTTATGCTTTTAAATTAATTAATTAATTAATTAATTAATTAAT 4287  
 DB 1080 TATTTTAAATGATATTTTATGCTTTTAAATTAATTAATTAATTAATTAATTAATTAAT 1133

RESULT 13  
 AC126053/c 230097 bp DNA linear HTG 02-JUL-2002  
 LOCUS Mus musculus chromosome UNK clone RP23-280L21, WORKING DRAFT  
 DEFINITION Mus musculus chromosome UNK clone RP23-280L21, WORKING DRAFT  
 AC126053  
 AC126053  
 AC126053 1 GI:21672255  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 230097)  
 McPherson, J.D. and Waterston, R.H.  
 TITLE The sequence of Mus musculus clone

JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 230097)  
 AUTHORS McPherson, J.D. and Waterston, R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-JUL-2002) Genome Sequencing Center, 4444 Forest Park  
 Parkway, St. Louis, MO 63108, USA  
 COMMENT

----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc/index.shtml  
 Contact: submissions@wustl.edu  
 Project Information  
 Center project name: M\_BA0280L21  
 ----- Summary Statistics -----  
 Sequencing vector: plasmid; 100%  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 225369 bases at least Q40  
 Consensus quality: 226638 bases at least Q30  
 Consensus quality: 227672 bases at least Q20

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 12 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1570: contig of 1570 bp in length  
 \* 1571 1670: gap of unknown length  
 \* 1671 5006: contig of 3336 bp in length  
 \* 5007 5106: gap of unknown length  
 \* 5107 10590: contig of 5484 bp in length  
 \* 10591 10690: gap of unknown length  
 \* 10691 16151: contig of 5461 bp in length  
 \* 16152 16251: gap of unknown length  
 \* 16252 24144: contig of 7893 bp in length  
 \* 24145 24244: gap of unknown length  
 \* 24245 37242: contig of 12998 bp in length  
 \* 37243 37343: gap of unknown length  
 \* 37343 50683: contig of 13341 bp in length  
 \* 50684 50783: gap of unknown length  
 \* 50784 68844: contig of 18060 bp in length  
 \* 68844 68943: gap of unknown length  
 \* 68943 94655: contig of 25712 bp in length  
 \* 94656 94755: gap of unknown length  
 \* 94756 136120: contig of 41365 bp in length  
 \* 136121 136220: gap of unknown length  
 \* 136221 176981: contig of 40761 bp in length  
 \* 176982 177081: gap of unknown length  
 \* 177082 230097: contig of 53016 bp in length.  
 Location/Qualifiers  
 1. 230097  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="UNK"  
 /clone="RP23-280L21"  
 1. 1570  
 /note="assembly\_name:Contig14"  
 1671. 5006  
 /note="assembly\_name:Contig16"  
 5107. 10590  
 /note="assembly\_name:Contig17"  
 10691. 16151  
 /note="assembly\_name:Contig18"  
 16252. 24144  
 /note="assembly\_name:Contig19"  
 24245. 37242

misc\_feature  
 /note="assembly\_name:Contig14"  
 1671. 5006  
 /note="assembly\_name:Contig16"  
 5107. 10590  
 /note="assembly\_name:Contig17"  
 10691. 16151  
 /note="assembly\_name:Contig18"  
 16252. 24144  
 /note="assembly\_name:Contig19"  
 24245. 37242  
 misc\_feature  
 /note="assembly\_name:Contig14"  
 1671. 5006  
 /note="assembly\_name:Contig16"  
 5107. 10590  
 /note="assembly\_name:Contig17"  
 10691. 16151  
 /note="assembly\_name:Contig18"  
 16252. 24144  
 /note="assembly\_name:Contig19"  
 24245. 37242  
 misc\_feature  
 /note="assembly\_name:Contig14"  
 1671. 5006  
 /note="assembly\_name:Contig16"  
 5107. 10590  
 /note="assembly\_name:Contig17"  
 10691. 16151  
 /note="assembly\_name:Contig18"  
 16252. 24144  
 /note="assembly\_name:Contig19"  
 24245. 37242







Query Match 17.4%; Score 883; DB 9; Length 1996;  
 Best Local Similarity 78.8%; Pred. No. 5,5e-192;  
 Matches 1286; Conservative 0; Mismatches 310; Indels 37; Gaps 18;

QY 3466 GAAATTTCCCAAGTTATTTGCTTCTTACGCTGCTACTGCTCCGAGACATCTACCTGGCT 3525  
 Db 132 GAAATTTCCCAAGTTATTTGCTTCTTACGCTGCTCCGAGAGCTCTCTCTCGCC 191  
 QY 3526 CCGGGGAGCTGATTTGATTCGGGAAAAAGAACCCAGTGGATGGTGGGAGAGAGACTG 3585  
 Db 192 CCGGTCAGCTGATTTGATTCGGGAAAAAGAACCCAGTGGATGGTGGGAGAGAGACTG 251  
 QY 3586 CAAGCTCGAGGAAAAAGCCGAGATAGGGTGGTTTCAGCAAAATTATGTCAAATCTTCTA 3645  
 Db 252 CAAGCACTGGGAAAAAGCCGAGATAGGGTGGTTTCAGCAAAATTATGTAAAGCTTCTA 311  
 QY 3646 AGCCCTGGGACGAGCAAAATCACTCCAGAGCCACCTTAAGTCAACACATTAAGCGCA 3705  
 Db 312 AGCCCTGGGACGAGCAAAATCACTCCAGAGCCACCTTAAGTCAACACATTAAGCGCA 371  
 QY 3706 GTGTGACAGGTGATTCGGGATGTAGATTAACCGCCAGAGCATGAGCACTAGCTTC 3765  
 Db 372 GTGTGACAGGTGATTCGGGATGTAGATTAACCGCCAGAGCATGAGCACTAGCTTC 431  
 QY 3766 AGCAAGGCGAGATCACTCAAGCTCTCAACAGAGAGACCCGAGCTGTGGAAAGAGAA 3825  
 Db 432 AACAGAGGCGAGATCACTCAAGCTCTCAACAGAGAGACCCGAGCTGTGGAAAGAGAA 491  
 QY 3826 GTAGTGGGCAAGTTGGGCTCTTCCATCAATTATGTAAAGCTGACACAGATGAGC 3885  
 Db 492 GTCAATGGACAAAGTGGGCTCTTCCATCAATTATGTAAAGCTGACACAGATGAGC 551  
 QY 3886 CCAGGCGAGATCACTCAAGCTCTCAACAGAGAGACCCGAGCTGTGGAAAGCTCAAG 3945  
 Db 552 CCAGGCGAGATCACTCAAGCTCTCAACAGAGAGACCCGAGCTGTGGAAAGCTCAAG 611  
 QY 3946 AGACCACTATCCCATATCACTGCGAGAGAGATGATGGAGATGAGGCTTGATCATGT 4005  
 Db 612 AGACCACTATCCCATATCACTGCGAGAGAGATGATGGAGATGAGGCTTGATCATGT 671  
 QY 4006 GACTTGAGCATATCACTGCTCTTCTGAGTAGAAGACTCACTGAGAG-CAGTTT 4064  
 Db 672 GACTTGAGCATATCACTGCTCTTCTGAGTAGAAGACTCACTGAGAGAGCAGTTT 731  
 QY 4065 ACCTCAATTTGACCTTGTGATGATGAAATGTCTGATCACTGCGTGGAGAGCGAG 4124  
 Db 732 ACCTCAATTTGACCTTGTGATGATGAAATGTCTGATCACTGCGAGAGATG 791  
 QY 4125 AAGC-AAATTTGAGAACTGACAGGGTGGTGGCTTTTGGGCTTCTTACTGACTC 4182  
 Db 792 GAGCAAAAATTAACAAAACACAGAGGTAGTGGCTTTTGGGCTTCTTACTGACTC 851  
 QY 4183 AGACTGAC-CGGCCCGGCTTCAACAGGGGCTTTCAATAGTTTAAATTTTAA 4241  
 Db 852 AAATTTGACTTTCCCAACCTTTCACAGGGGCTTTCAATAGTTTAAATTTTAA 911  
 QY 4242 TGATGATTTAGCTTTTAAATAATTCATCAATTAATTTCTTGGCTATTTGGTT 4301  
 Db 912 TATATATTTAGCTTTTAAATAATTCATCAATTAATTTCTTGGCTATTTGGTT 971  
 QY 4302 TACAAAAACACCACTATCAAGAGTCTGCTGGGAGAGATTAATGCTGTTCCGG 4361  
 Db 972 TGCAAAAAGACCACTATCAAGAGTCTGCTGGGAGAGATTAATGCTGTTCCAA 1026  
 QY 4362 CGTACCTTAACTGAGAGCTTGTGATCTTGGCGGTTTGTCCAGTGTCCCAACACT 4421  
 Db 1027 ATGTCCTAATAATCTGAGACTGTGATTTTTCATTTTTCAGAGTGTACCAACTAAT 1086  
 QY 4422 TGATGATTTGGGCTGTTCCCTGCGGTAGAGACAGAGAGATGGGTACTCTGT 4477  
 Db 1087 TGATGATTTGGGCTGTTCCCTGCGGTAGAGACAGAGAGATGGGTACTCTGT 1146

QY 4478 TTTGAAATGTGTATGTAGACTGAGCTGACTATGAGAGGGTATGTCTGTCTGACC 4537  
 Db 1147 TTTTAAAGCGTATA-GATATGACCCAAATTAAGAGAGGTTGTGTCTGTGTGTGT 1205  
 QY 4538 ATTCAGGTATCC-TGTGGCGATGTACCATCTGTACCGAAGAGTACT-CTTCCCTCAT 4595  
 Db 1206 ATTCAGGTATCTTGTGAGATGTATATACCTGTACATTAAGAAATTTAGTTCTTCCA 1265  
 QY 4596 GGTAAACCCACACCGGTATGACAGTGTCTCATCTACT-GCATTCATTTTACTTTTGCAC 4653  
 Db 1266 TGCAAGACTTATTAATCTTGTATGATGATGCTTAATATATATGATTAATTTATTTGAC 1325  
 QY 4654 AGTACCTTTGAGCCACTGAGAGACAC-CCATGTTCCGTTTGGTCTCAGATGTACT 4712  
 Db 1326 AGTACCTTTGAGCCACTGAGAGACACCTGTGTTTGTGTTGGTCTCAGATTTATCT 1385  
 QY 4713 AGTTGTCCGCTGTTTGTATTTTCAATGTGACATGTCTGACACATTAAGTAG 4772  
 Db 1386 GATTGAGTTGATGTTGTTGTTGGGTTTATATTTCTGCTGTTTGCATGACATTAATCA 1445  
 QY 4773 TAAAGCGCAACTGCGGAGGCTTACATCATGATGATGACCCAGCTC-----TTAGTCTC 4827  
 Db 1446 GTAAGCAAC-ACCACTAGAGCTGTACGATCAAGATATCAAGTCTCTTTTATGCTC 1504  
 QY 4828 TGTATCGTAAG-TTATTTCCAGTTGCTTTTATGAA-----TATCTTGAACAATAT 4881  
 Db 1505 TGTATCAATGAATTTTATTTCCAGTTACTTTTCAATGATGATGATGATTTTAAACAATAT 1564  
 QY 4882 CTCTTGAACAAGAAATGATATGAAAGTCCCTGCAATTAATTTCCAGTGTTCAT 4941  
 Db 1565 TTTCTTGAACAAGAAATGATATGAAAGTCCCTGCAATTAATTT-CCAATGTTTACAT 1623  
 QY 4942 TTTTAACTAAGCTGTGGGGTGTCTACAGATTAATGAAATGGCCCTGCTGCTCGT 5001  
 Db 1624 TTTTAACTAAGCTGT-AGATTTCTACAGATTAATGAAATGAGACTCATGCTCGTT 1682  
 QY 5002 TGTGTATTAATCTGTCTGTATGCTGAAGCGGTGT--CCTTAGATATTAATTTGGAAT 5058  
 Db 1683 TGTGTG-TAAGATATGCTGTGTGATGAAGCCTGTTGTCTTTTAAACATATTTGGAAGC 1741  
 QY 5059 CGGAAAGGAATT 5071  
 Db 1742 TCTCAATTAATAT 1754

RESULT 15  
 AC106760/c 78190 bp DNA linear PRI 27-MAR-2002  
 LOCUS  
 DEFINITION Homo sapiens chromosome 5 clone RP11-215G15, complete sequence.  
 AC106760  
 VERSION  
 AC106760.2 GI:19747162  
 KEYWORDS  
 HTG.  
 SOURCE  
 Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1 (bases 1 to 78190)  
 DOE Joint Genome Institute and Stanford Human Genome Center.  
 TITLE  
 Direct Submission  
 JOURNAL  
 Unpublished  
 2 (bases 1 to 78190)  
 DOE Joint Genome Institute.  
 TITLE  
 Direct Submission  
 JOURNAL  
 Submitted (12-JUN-2002) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 3 (bases 1 to 78190)  
 DOE Joint Genome Institute and Stanford Human Genome Center.  
 TITLE  
 Direct Submission  
 JOURNAL  
 Submitted (27-MAR-2002) DOE Joint Genome Institute, 2800 Mitchell  
 Drive, Walnut Creek, CA 94598, USA  
 On Mar 27, 2002 this sequence version replaced gi:18139310.  
 COMMENT  
 Draft Sequence Produced by DOE Joint Genome Institute  
 www.jgi.doe.gov

Finishing Completed at Stanford Human Genome Center  
 www.shgc.stanford.edu  
 Quality: Phrap Quality >=40 99.7% of Sequence;  
 Estimated Total Number of Errors is 0.3.  
 NOTE: This insert is not the entire sequence of the clone (entire  
 sequence is 179,4kb). It is clipped at the overlap with AC012629.  
 The number of bases overlapped is 5036.

## FEATURES

source

1. 78190

/organism="Homo sapiens"

/db xref="taxon:9606"

/chromosome="5"

/clone="RP11-215G15"

BASE COUNT 19148 a 18124 c 19405 g 21513 t

## ORIGIN

Query Match 16.0%; Score 813.6; DB 9; Length 78190;  
 Best Local Similarity 88.6%; Pred. No. 7,6e-176;  
 Matches 927; Conservative 0; Mismatches 114; Indels 5; Gaps 4;

QY 3225 GCGCCGCAAGCCGACCATTCGCCGAGAGAGTTATTGCGCATGTACACATACGAGAGTTG 3284  
 Db 73571 GACTGCTAACCCAGTCGTTTCGGAGAGATTATTCCTATGTACACTTACCAAGAGTTG 73512  
 QY 3285 TGAGCAAGAGATTATTAACCTTTGAGCAAGGGAGATGTGATTGTGTACCAAGAAAGATGG 3344  
 Db 73511 TGAGCAAGAGATTATTAACCTTTGAGCAAGAGAGATGTGATTGTGTACCAAG-AAAGATGG 73453  
 QY 3345 TGACTGGTGACGGGAAAGGCTGAGCAAGTCGGAAGTCTTCCCTTCTAATCTATGTGTGAG 3404  
 Db 73452 TGACCGGTGACAGAGAAACAGTGGGCGCAAGCGCGGTCTTCCCTTCTAATCTATGTGTGAG 73393  
 QY 3405 GCTTAAAGATTGACAGGGGCTCTGAACTGCT-GGGAAGAACAGGAGTTTAAAGAAAGAAAC 3463  
 Db 73392 GCTTAAAGATTGACAGGGGCTCTGAACTGCTGAGAAATGAGGAGTTTAAAGAAAGAAAC 73333  
 QY 3464 CTGAATTTGCCAGGTATTGCTTCTACCGCTCTACTGGTCCCGAACAACCTCACCCCTGG 3523  
 Db 73332 CTGAATTTGCCAGGTATTGCTTCTACACCGCCACGCGCCCGAGCAGCTCCTCTCG 73273  
 QY 3524 CTCTGGGCGAGCTGATTCTATCCGGAAGAAAGCCAGGTGATGTGGGAAGGAAGAAC 3583  
 Db 73272 CCCTGGTCACTGATTCTGATTTGATCTGAAAAAGAACCAAGGTGATGTGGGAAGGAAGAAC 73213  
 QY 3584 TGCAGCTCAGAGGAAAAAGCCAGATAGGTGCTTCCAGCAATTAATGTCAAACTTC 3643  
 Db 73212 TGCAGCTCAGGTGGAAAAAGCCAGATAGGTGCTTCCAGCAATTAATGTCAAACTTC 73153  
 QY 3644 TAAAGCCCGGAGCAAGCAAAATCAACCCCACTAGCTTACCAAGACCGACAGTGCAGCCAG 3703  
 Db 73152 TAAAGCCCGGAGCAAGCAAAATCACTCCACAGGCCACTAAGTCAACAGCATTTAGCGG 73093  
 QY 3704 CAGTGTCCAGGTGATTCGGAGATGATGATTAACCGCCGAGAACGATGAGAGACTAGCCCT 3763  
 Db 73092 CAGTGTCCAGGTGATTCGGAGATGATGATTAACCGCCGAGAACGATGAGAGACTAGCCCT 73033  
 QY 3764 TCAGCAAGGCGCAGATCATCAAGCTCTCAACAAGAGAGACCCGGACTGGTGGAAAGAGAG 3823  
 Db 73032 TCAGCAAGGCGCAGATCATCAAGCTCTCAACAAGAGAGAGACCCGGACTGGTGGAAAGAGAG 72973  
 QY 3824 AAGTCACTGGGCAAGTTGGGCTCTTCCATCTCAATTAATGTAAGCTGACCAAGACATGG 3883  
 Db 72972 AAGTCACTGGGCAAGTTGGGCTCTTCCATCTCAATTAATGTAAGCTGACCAAGACATGG 72913  
 QY 3884 ACCCGAGCGCAGCAATGATATATGTTGTCATCCCGCCCTCAAGGCTTGAAGTCCCTCAA 3943  
 Db 72912 ACCCGAGCGCAGCAATGATATATGTTGTCATCCCGCCCTCAAGGCTTGAAGTCCCTCAA 72853  
 QY 3944 AGAGACCACTATTCATATCACTGCCAGAGGAGATGATGGAGATGACAGCTTGTATCAT 4003  
 Db 72852 AGAGACCACTATTCATATCACTGCCAGAGGAGATGATGGAGATGACAGCTTGTATCAT 72793  
 QY 4004 GTGACTTGACAGCATGATCAGCTCTGCTTCTAGTAGAAGAACTCACTGCAGAGCAGTT 4063

Db 72792 GTGACTTGACAGCATGATCAGCTCTGCTTCTAGTAGAAGAACTCACTGCAGAGCAGTT 72733  
 QY 4064 TACCTCATTTGACCTTATGATGATGATGATGAAATGTCTGAGTCACTGCTGCAGAGGCA 4123  
 Db 72732 TACCTCATTTGACCTTATGATGATGATGATGAAATGTCTGAGTCACTGCTGCAGAGGCA 72673  
 QY 4124 GAAGC-AAATGTGAGAACTGCACAGGAGTGTGGTCTTTGGGCTTCTTACGTCACCT 4181  
 Db 72672 GGAGCAAAAAATTAACAAAAATACACAGGAGAGGAGTGTGGTCTTTGTGGCTTCTTACGTCACCT 72613  
 QY 4182 CAGACTGAC-CGCGCCCGCCTTCACACAGGCGCTTTCATATGTTTAAAGATTATTTTAA 4240  
 Db 72612 CAATTTGACTTTCCCGCAGCTTTGCAACAGGTGCTTCAATATGTTTAAAGATTATTTTAA 72553  
 QY 4241 ATGTATTTTACCTTTTAAATAAA 4266  
 Db 72552 ATATATTTTACCTTTTAAATAAA 72527

Search completed: March 14, 2003, 08:16:31  
 Job time : 12832 secs



XX New nucleic acid encoding Ese1 and 2 proteins, involved in regulation  
PT of endocytosis, used e.g. for treating cancer or preventing viral  
PT infection -  
XX  
XX  
PS Claim 6; Page 38-40; 99pp; English.  
XX

CC The present sequence encodes mouse Ees1. The present invention  
CC specifically describes mammalian Ees1 and 2 proteins (I) and their splice  
CC variants (Ese = EH-domain and SH3-domain regulator of endocytosis). (I)  
CC are involved in regulation of clathrin-mediated endocytosis (as a complex  
CC with Eps15 protein), vesicular trafficking and actin cytoskeleton.  
CC Generally (I) or its (ant)agonists, mimetics, fragments and inactive  
CC mutants); (I)-specific antibodies (Ab); sequences antisense to the (I)  
CC polynucleotide; agents that downregulate expression of Ees genes or  
CC antagonists of an Ees binding partner are used to treat diseases  
CC associated with undesirable endocytosis and resulting changes in cellular  
CC function. Particularly overexpression of Ees1 is used to block  
CC clathrin-mediated endocytosis in vivo or in cell cultures, while  
CC administration of (I) is used to promote endocytosis of selected cells.  
CC (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of  
CC cells that can be stimulated to proliferate by a growth factor receptor;  
CC and similar compounds (also inactive Ese mutants) can be used to prevent  
CC viral infection. Endocytosis may also be regulated, in vivo or in cell  
CC cultures, by forming an Ese-Eps15 complex, then binding dynamin to the  
CC complex. Generally conditions that can be treated include cancer;  
CC abnormal cell division or mitration; viral infection; or abnormal  
CC receptor signalling, tissue development or synaptic transmission.  
XX  
50 Sequence 5082 BP, 1363 A, 1253 C, 1391 G, 1075 T, 0 other;

Query Match	99.5%;	Score 5060;	DB 21;	Length 5082;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 5082;	Conservative 0;	Mismatches 0;	Indels 2;	Gaps 2

OY	1	CGGCACGAGAGAGATGAGAGCGCGCGGAGAGGGCTGCGAGCTTGTTGCTCCGTAGTACG	60
Db	1	CGGCACGAGAGAGATGAGAGCGCGCGGAGAGGGCTGCGAGCTTGTTGCTCCGTAGTACG	60
OY	61	CGCGCTCGCAAGGAGACATCCGACGCGGCTCCGGACCGCCCGGAGGCAAGCAGCGCGG	120
Db	61	CGCGCTCGCAAGGAGACATCCGACGCGGCTCCGGACCGCCCGGAGGCAAGCAGCGCGG	120
OY	121	CGGGCGGGGATGGTGTGCGCGCGGTGCGGACTCGGCGTTCTCGCGCGGGGTGGGGCTGC	180
Db	121	CGGGCGGGGATGGTGTGCGCGCGGTGCGGACTCGGCGTTCTCGCGCGGGGTGGGGCTGC	180
OY	181	ACTGATTTGTGAGAGGGGCGCGCGCAGCACCGCCCGAGATGAGAGCGCTGCATCAGCAA	240
Db	181	ACTGATTTGTGAGAGGGGCGCGCGCAGCACCGCCCGAGATGAGAGCGCTGCATCAGCAA	240
OY	241	GGTGAACGTAATAGAACCAATGGCTCAGTTTCCACACTTTTGGTGTAGCTCGATGTC	300
Db	241	GGTGAACGTAATAGAACCAATGGCTCAGTTTCCACACTTTTGGTGTAGCTCGATGTC	300
OY	301	TGGGCATTAAGTGTGAGAGGAAAGGGCGCAAGCATGACACAGTTCCTTAGCCGTGAAGCGG	360
Db	301	TGGGCATTAAGTGTGAGAGGAAAGGGCGCAAGCATGACACAGTTCCTTAGCCGTGAAGCGG	360
OY	361	ATAGCGGGATTTATTACTGTGTGATCAAGCGAGGAACCTTTTTTCCATCTGGGTTACCT	420
Db	361	ATAGCGGGATTTATTACTGTGTGATCAAGCGAGGAACCTTTTTTCCATCTGGGTTACCT	420
OY	421	CAGCGTGTATGACCAAAATATGGGCGGTAGGGGACATGAATAGCATGAAGAGATGAT	480
Db	421	CAGCGTGTATGACCAAAATATGGGCGGTAGGGGACATGAATAGCATGAAGAGATGAT	480
OY	481	CAAGTGAATTTTCATAGACCATGAAGCTTATCAACTGAAGCTACAGAGTATCAGCTC	540
Db	481	CAAGTGAATTTTCATAGACCATGAAGCTTATCAACTGAAGCTACAGAGTATCAGCTC	540
OY	541	CCCTCCACACTTCCCCCTCTCATGAACAAGCAACAGTGAGTATTTCCAGTGCACCAAGA	600
Db	541	CCCTCCACACTTCCCCCTCTCATGAACAAGCAACAGTGAGTATTTCCAGTGCACCAAGA	600

Db	541	CCCTCCACACTTCCCTCGTCATGAAACAGCAACAGTG3GCTAATTTCCAGTGCACCCAGA	600
Qy	601	TTTGGTATAGAGGGAGTTGCTAGCATGCCACCACTCAAGCTGTGTCTCTGTGCCAATG	660
Db	601	TTTGGTATAGAGGGAGTTGCTAGCATGCCACCACTCAAGCTGTGTCTCTGTGCCAATG	660
Qy	661	GAGCTCATTCAGTAGTTGTGGAATGTCCACCCCTTAAGTATCTTGTGCTCCACAGACA	720
Db	661	GAGCTCATTCAGTAGTTGTGGAATGTCCACCCCTTAAGTATCTTGTGCTCCACAGACA	720
Qy	721	GTGCTCTCCCTGGCTTAACGGGGCTCTCCCGTATACAGCTCTGTGCTGCTTGGCGAT	780
Db	721	GTGCTCTCCCTGGCTTAACGGGGCTCTCCCGTATACAGCTCTGTGCTTGGCGAT	780
Qy	781	CTTGACGCACATGCGCAAGAGTCTCTTCAGAGATCTGTGTCACAGGGCACAAATTA	840
Db	781	CTTGACGCACATGCGCAAGAGTCTCTTCAGAGATCTGTGTCACAGGGCACAAATTA	840
Qy	841	AACACTAAGTTACAGAAAGGACCAATCATTTGATGTGCGAGCGCCCTTCACAGCAGAA	900
Db	841	AACACTAAGTTACAGAAAGGACCAATCATTTGATGTGCGAGCGCCCTTCACAGCAGAA	900
Qy	901	TGGGCTGTGCGCTACGTATCAAGGCTGAAATACAGGCAATTTTCAACAGCCAGCAAA	960
Db	901	TGGGCTGTGCGCTACGTATCAAGGCTGAAATACAGGCAATTTTCAACAGCCAGCAAA	960
Qy	961	ACTATGAGTGGACACTTACAGGCTCCCAAGGCAAGCACTATTTCTGATGCAATCAAGTTTA	1020
Db	961	ACTATGAGTGGACACTTACAGGCTCCCAAGGCAAGCACTATTTCTGATGCAATCAAGTTTA	1020
Qy	1021	CCCCAGGCTACGTGCGTTCAATATGGAATCTTTCTGAATTTGATTCAGATGGAATACTC	1080
Db	1021	CCCCAGGCTACGTGCGTTCAATATGGAATCTTTCTGAATTTGATTCAGATGGAATACTC	1080
Qy	1081	ACTGACAGAGATTTTATCTACGCTATGSCACTAATGATGTTGCCATGCTGTGACGCA	1140
Db	1081	ACTGACAGAGATTTTATCTACGCTATGSCACTAATGATGTTGCCATGCTGTGACGCA	1140
Qy	1141	CTGCGGCCGCTCTGCTCCAGAAATACATCCCTCTTCTTCAAGAAAGTTGCTCGGC	1200
Db	1141	CTGCGGCCGCTCTGCTCCAGAAATACATCCCTCTTCTTCAAGAAAGTTGCTCGGC	1200
Qy	1201	AGTGGAGTGTCCGTCATTAACCTCTTCTTCGTGAGATCAAGGCTGCTGAGAGCCGTG	1260
Db	1201	AGTGGAGTGTCCGTCATTAACCTCTTCTTCGTGAGATCAAGGCTGCTGAGAGCCGTG	1260
Qy	1261	TCAAGAGATGACAGCAAGCCAGAGAAAGAACTGCTGTGACATTTGAAGATTAAGACGG	1320
Db	1261	TCAAGAGATGACAGCAAGCCAGAGAAAGAACTGCTGTGACATTTGAAGATTAAGACGG	1320
Qy	1321	GAGAACTTTCAGGCGAGCACTGTGAGACTGTGAGAAAGCGCGCAAGCGCTCTTGGAGAG	1380
Db	1321	GAGAACTTTCAGGCGAGCACTGTGAGACTGTGAGAAAGCGCGCAAGCGCTCTTGGAGAG	1380
Qy	1381	CAGCGCAAAAGCAGAGCCGTTGTGCTTCAGCTTGAAGCGCGGACAGAGAGGAAAGAG	1440
Db	1381	CAGCGCAAAAGCAGAGCGGTTGTGCTTCAGCTTGAAGCGCGGACAGAGAGGAAAGAG	1440
Qy	1441	CGGAGCGCCAGAGAGCAGAGAGGCCAAGCGCGCAGCTGTGAAGCTTGAAGACAGCTTGAAGAG	1500
Db	1441	CGGAGCGCCAGAGAGCAGAGAGGCCAAGCGCGCAGCTGTGAAGCTTGAAGACAGCTTGAAGAG	1500
Qy	1501	CAGCGGAGCTGTGAGCGGCGCAGAGAGAGAGAGAGAGAGAGAGATGAGAGCGCGCAG	1560
Db	1501	CAGCGGAGCTGTGAGCGGCGCAGAGAGAGAGAGAGAGAGAGAGATGAGAGCGCGCAG	1560
Qy	1561	GCCGCAGAAACGGGAATCTGAAAGCAGCGACAACTTGAATGGAAACGAAACCGAGACAG	1620
Db	1561	GCCGCAGAAACGGGAATCTGAAAGCAGCGACCAACTTGAATGGAAACGAAACCGAGACAG	1620
Qy	1621	GAACTCTGTAATCAGAGGAACAAGAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1680
Db	1621	GAACTCTGTAATCAGAGGAACAAGAGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1680

QY 1681 AAGACTCTGGAGTTGAGTTAGAACTGAAATGACAAAAGCATGCTAGAAAGGAAA 1740  
DB 1681 AAGACTCTGGAGTTGAGTTAGAACTGAAATGACAAAAGCATGCTAGAAAGGAAA 1740  
QY 1741 CTTGAGGATATCAGTGTGCTGCACTGCGAACCCAGAGGAGGAAATTTGAGAGCAGCAACAG 1800  
DB 1741 CTTGAGGATATCAGTGTGCTGCACTGCGAACCCAGAGGAGGAAATTTGAGAGCAGCAACAG 1800  
QY 1801 TCTAGAGAGTAAAGATGTGTAATACCCACTTACAGAGAGTGTGAGGAAATCTGAG 1860  
DB 1801 TCTAGAGAGTAAAGATGTGTAATACCCACTTACAGAGAGTGTGAGGAAATCTGAG 1860  
QY 1861 CAATCTTGGAGAGCTTATTCAGAGAAAACAGATACAGTACAGTAAACAAAGTC 1920  
DB 1861 CAATCTTGGAGAGCTTATTCAGAGAAAACAGATACAGTACAGTAAACAAAGTC 1920  
QY 1921 CAGCAGAACAGTTTGATAGAGACTGCTTCTTACCTTCAAAAAGCCTTTGAGAACAG 1980  
DB 1921 CAGCAGAACAGTTTGATAGAGACTGCTTCTTACCTTCAAAAAGCCTTTGAGAACAG 1980  
QY 1981 GAGCTGCGCCGAGAGAGCTCCGAGAGCAGTGGAGAGTGGAGAGAGACAGGTTCA 2040  
DB 1981 GAGCTGCGCCGAGAGAGCTCCGAGAGCAGTGGAGAGTGGAGAGAGACAGGTTCA 2040  
QY 2041 AAGCTCAGAGAGATGATGTTTCAACAAACAGCTGAGAACTGAGAGAGATACATAGC 2100  
DB 2041 AAGCTCAGAGAGATGATGTTTCAACAAACAGCTGAGAACTGAGAGAGATACATAGC 2100  
QY 2101 AAAACAGAACTCCAGAGAGCAGAGGCTCTGAGAGCAGCGGAGCTGAGAGCAAGAGCAG 2160  
DB 2101 AAAACAGAACTCCAGAGAGCAGAGGCTCTGAGAGCAGCGGAGCTGAGAGCAAGAGCAG 2160  
QY 2161 GAGAGGAAAGCCTGAGATTTAGAGAAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 2220  
DB 2161 GAGAGGAAAGCCTGAGATTTAGAGAAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 2220  
QY 2221 AAGGACAAAGAGAGCTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280  
DB 2221 AAGGACAAAGAGAGCTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280  
QY 2281 CACGAG 2340  
DB 2281 CACGAG 2340  
QY 2341 GCCAAGCCGGAATTCAG 2400  
DB 2341 GCCAAGCCGGAATTCAG 2400  
QY 2401 AAGCTGCGCAACAG 2460  
DB 2401 AAGCTGCGCAACAG 2460  
QY 2461 CAGAGAGAGTAAAGTGTATATTTACCGAGCGCTGTATCCCTTTGATTCAGAGAGTCA 2520  
DB 2461 CAGAGAGAGTAAAGTGTATATTTACCGAGCGCTGTATCCCTTTGATTCAGAGAGTCA 2520  
QY 2521 GATGAGATTCACATCCAG 2580  
DB 2521 GATGAGATTCACATCCAG 2580  
QY 2581 CCAGAGATGCTGAG 2640  
DB 2581 CCAGAGATGCTGAG 2640  
QY 2641 GAAAAAGATTCAG 2700  
DB 2641 GAAAAAGATTCAG 2700  
QY 2701 GCGCCCTGCGCCCAAG 2760  
DB 2701 GCGCCCTGCGCCCAAG 2760

QY 2761 GAGCCCTTCACACACCCCAACCACTGGGCGAGCTTCACTTCCAGCTGCGCCAGCAGCTCA 2820  
DB 2761 GAGCCCTTCACACACCCCAACCACTGGGCGAGCTTCACTTCCAGCTGCGCCAGCAGCTCA 2820  
QY 2821 AACGAGAAAG 2880  
DB 2821 AACGAGAAAG 2880  
QY 2881 CCTAGTGTGCGCAGTTACGAGAGAGATCAGCTTATCCCGAGCAGAGCAGTGGCTCC 2940  
DB 2881 CCTAGTGTGCGCAGTTACGAGAGAGATCAGCTTATCCCGAGCAGAGCAGTGGCTCC 2940  
QY 2941 TCCCATCTCCGCTCTGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3000  
DB 2941 TCCCATCTCCGCTCTGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3000  
QY 3001 TATCCCTGAG 3060  
DB 3001 TATCCCTGAG 3060  
QY 3061 GTTCTGGAACAGCAAG 3120  
DB 3061 GTTCTGGAACAGCAAG 3120  
QY 3121 CCAAGCTTACGAG 3180  
DB 3121 CCAAGCTTACGAG 3180  
QY 3181 GAGCCCTTACGAG 3240  
DB 3181 GAGCCCTTACGAG 3240  
QY 3241 ATTCCCGAG 3300  
DB 3241 ATTCCCGAG 3300  
QY 3301 ACCCTTCAG 3360  
DB 3301 ACCCTTCAG 3360  
QY 3361 ACGGTGGCGAG 3420  
DB 3361 ACGGTGGCGAG 3420  
QY 3421 GAGCTGGAACAG 3480  
DB 3421 GAGCTGGAACAG 3480  
QY 3481 ATTGCTTCTACAGCTGCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3540  
DB 3481 ATTGCTTCTACAGCTGCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3540  
QY 3541 CTGATCCGAG 3600  
DB 3541 CTGATCCGAG 3600  
QY 3601 AAGCCGAGATGAGTGTGTTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660  
DB 3601 AAGCCGAGATGAGTGTGTTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660  
QY 3661 AAAATCACCCCACTGAG 3720  
DB 3661 AAAATCACCCCACTGAG 3720  
QY 3721 GGGATGATAGATTAACCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3780  
DB 3721 GGGATGATAGATTAACCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3780  
QY 3781 ATCAAGCTCTCAACAG 3840  
DB 3781 ATCAAGCTCTCAACAG 3840  
QY 3841 GGGCTTCTCCATCAATTAATTAAGTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3900  
DB 3841 GGGCTTCTCCATCAATTAATTAAGTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3900

Db 3840 GGGGCTTCCCATCCAAATTATGTAAGCTGACCAAGACATGAGCCAGCCAGCCAAATGA 3899  
 Qy 3901 ATCATATGTTGTCATCCCGCCCTGAGGCTGGAAGTCTTAAAGAGACCACATATCCA 3960  
 Db 3900 ATCATATGTTGTCATCCCGCCCTGAGGCTGGAAGTCTTAAAGAGACCACATATCCA 3959  
 Qy 3961 TATCACTGCCAGAGAGATGATGGAGATGAGAGCTTGATCATGTGACTGACAGATGAT 4020  
 Db 3960 TATCACTGCCAGAGAGATGATGGAGATGAGAGCTTGATCATGTGACTGACAGATGAT 4019  
 Qy 4021 CACCTACTGCTTCTGATAGAAAGACCTGAGAGAGAGATTTACTCTATTTACCTTA 4080  
 Db 4020 CACCTACTGCTTCTGATAGAAAGACCTGAGAGAGAGATTTACTCTATTTACCTTA 4079  
 Qy 4081 GTTGATGTGATCGAAATGTCAGTCACTGCGAGAGAGAGAAATTCAGAAAC 4140  
 Db 4080 GTTGATGTGATCGAAATGTCAGTCACTGCGAGAGAGAGAAATTCAGAAAC 4139  
 Qy 4141 TGCACAGGGGTGGGTCTTTTGGGGCTTCTAGTCACTGAGACTGACCGGCCGCC 4200  
 Db 4140 TGCACAGGGGTGGGTCTTTTGGGGCTTCTAGTCACTGAGACTGACCGGCCGCC 4199  
 Qy 4201 TTTCAAGGGGCTTTCATATGTTTAAAGATTTTAAATGTTTAAAGCTTTA 4260  
 Db 4200 TTTCAAGGGGCTTTCATATGTTTAAAGATTTTAAATGTTTAAAGCTTTA 4259  
 Qy 4261 ATAAATCTCAATCAATTAATCTTCTGCTATTTTGGTTTAAAGAGAGAGAGATTC 4320  
 Db 4260 ATAAATCTCAATCAATTAATCTTCTGCTATTTTGGTTTAAAGAGAGAGATTC 4319  
 Qy 4321 AAGAGAGGCTGCTGCGGAGAGATTTAAATGCTGTTCGGGGCTGACCTGTAACAGAGAGC 4380  
 Db 4320 AAGAGAGGCTGCTGCGGAGAGATTTAAATGCTGTTCGGGGCTGACCTGTAACAGAGAGC 4379  
 Qy 4381 TTGCTGTAACCTTTGCGGTTTGTCCAGTGTTCACCAACCAATTTGTGTGGGGCTGT 4440  
 Db 4380 TTGCTGTAACCTTTGCGGTTTGTCCAGTGTTCACCAACCAATTTGTGTGGGGCTGT 4439  
 Qy 4441 CCTGCGGTAGAGAGAGAGATGGGTGTAAGCTGTTTGAAGATGTGATGAGACTG 4500  
 Db 4440 CCTGCGGTAGAGAGAGAGATGGGTGTAAGCTGTTTGAAGATGTGATGAGACTG 4499  
 Qy 4501 AGCCTGATATGAAAGGGGTATGCTGTGACCATCAAGTGTGCTGGGGCTGAG 4560  
 Db 4500 AGCCTGATATGAAAGGGGTATGCTGTGACCATCAAGTGTGCTGGGGCTGAG 4559  
 Qy 4561 TACCATCTGTAACGAGAGAGTCTTCTCAATGCTAAACCCAGCCAGCTGATCAGT 4620  
 Db 4560 TACCATCTGTAACGAGAGAGTCTTCTCAATGCTAAACCCAGCCAGCTGATCAGT 4619  
 Qy 4621 GCTCTCATCTACCTGATTTTACTTTGACAGTACCTTTGAGCCACTGAGAGAGC 4680  
 Db 4620 GCTCTCATCTACCTGATTTTACTTTGACAGTACCTTTGAGCCACTGAGAGAGC 4678  
 Qy 4681 ACCCATGTTCCGTTGGTGTCAAGATGATCTAGTTGACCGGCTGTTTATTTT 4740  
 Db 4679 ACCCATGTTCCGTTGGTGTCAAGATGATCTAGTTGACCGGCTGTTTATTTT 4738  
 Qy 4741 TCAATCTGGCATGTTTCAACCACTAAGTGAAGCGCCAACTGCCAGGCGGTTACG 4800  
 Db 4739 TCAATCTGGCATGTTTCAACCACTAAGTGAAGCGCCAACTGCCAGGCGGTTACG 4798  
 Qy 4801 ATATCAGTACCCAGCGCTTACGCTGTTAGAGTGAAGTTTATTCAGTTGCTTTTAT 4860  
 Db 4799 ATATCAGTACCCAGCGCTTACGCTGTTAGAGTGAAGTTTATTCAGTTGCTTTTAT 4858  
 Qy 4861 GGAATATCTTGAACAGTAATCTTCTGACAGAGAAAGATGATAGAGTCTCCCTGCA 4920  
 Db 4859 GGAATATCTTGAACAGTAATCTTCTGACAGAGAAAGATGATAGAGTCTCCCTGCA 4918  
 Qy 4921 TTAATTTCCAGGTTTACATTTTAACTAGACTGTGGGGGTTGCTACGATTTAATAG 4980

Db 4919 TTAATTTCCAGTGTTTACATTTTAACTAGACTGTGGGGTGTCTACAGATTAATAG 4978  
 Qy 4981 AATGGCGCTCTGCGCTGTGTGTTAACTTTGCTGTAGAGTGAAGCCGTGTGCT 5040  
 Db 4979 AATGGCGCTCTGCGCTGTGTGTTAACTTTGCTGTAGAGTGAAGCCGTGTGCT 5038  
 Qy 5041 TAGATATTAGTTGGAAGTCCGAGAGAGAAATTCGATATCAAGCTT 5084  
 Db 5039 TAGATATTAGTTGGAAGTCCGAGAGAGAAATTCGATATCAAGCTT 5082  
 RESULT 2  
 ID AA239024 standard; cDNA; 5738 BP.  
 AC AA239024;  
 XX  
 DT 28-FEB-2000 (first entry)  
 XX  
 DE Mouse Esei1 cDNA sequence.  
 XX  
 KW Mouse; murine; Esei1; Esei2; endocytosis; vesicular trafficking;  
 KW regulation; actin cytoskeleton; detection; cancer; infection;  
 KW EH-domain and SH3-domain regulator of endocytosis; anticancer;  
 KW antiproliferative; antiviral; ss.  
 XX  
 OS Mus sp.  
 XX  
 PN W09955728-A2.  
 XX  
 PD 04-NOV-1999.  
 XX  
 PF 27-APR-1999; 99WO-CA00375.  
 XX  
 PR 27-APR-1998; 98CA-2230201.  
 XX  
 PR 05-FEB-1999; 99US-0118739.  
 XX  
 PA (HSCR-) HSC RES & DEV LP.  
 XX  
 PI Egan SE, Wang W, Sengar A;  
 DR WPI, 2000-052802/04.  
 DR P-PSDB; AAY57449.  
 XX  
 PT New nucleic acid encoding Esei1 and 2 proteins, involved in regulation  
 PT of endocytosis, used e.g. for treating cancer or preventing viral  
 PT infection  
 XX  
 PS Claim 6; Page 56-59; 99p; English.  
 XX  
 CC The present invention specifically describes mammalian Esei1 and 2  
 CC proteins (I) and their splice variants (Ese = EH-domain and SH3-domain  
 CC regulator of endocytosis). (I) are involved in regulation of clathrin-  
 CC mediated endocytosis (as a complex with Esi15 protein), vesicular  
 CC trafficking and actin cytoskeleton. Generally (I) (or its (ant)agonists,  
 CC mimetics, fragments and inactive mutants); (I)-specific antibodies (Ab);  
 CC sequences antisense to the (I) polynucleotide; agents that downregulate  
 CC expression of Esei genes or antagonists of an Esei binding partner are  
 CC used to treat diseases associated with unstable endocytosis and  
 CC resulting changes in cellular function. Particularly overexpression of  
 CC Esei1 is used to block clathrin-mediated endocytosis in vivo or in cell  
 CC cultures, while administration of (I) is used to promote endocytosis of  
 CC selected cells. (Ant)agonists of (I) or Ab are used to suppress abnormal  
 CC proliferation of cells that can be stimulated to proliferate by a growth  
 CC factor receptor; and similar compounds (also inactive Esei mutants) can be  
 CC used to prevent viral infection. Endocytosis may also be regulated, in  
 CC vivo or in cell cultures, by forming an Esei-Esi15 complex, then binding  
 CC dynamin to the complex. Generally conditions that can be treated include  
 CC cancer; abnormal cell division or migration; viral infection; or abnormal  
 CC receptor signalling; tissue development or synaptic transmission. The  
 CC present sequence represents mouse Esei1 cDNA sequence.  
 XX  
 SQ Sequence 5738 BP; 1556 A; 1456 C; 1613 G; 1113 T; 0 other;







Db 2101 CAAAGCACTCCAGAGCAGAGGTCCCTGGAGCAGCGCGACTGAGAGCAAAAGCA 2160  
 Qy 2160 GAGAGGAAGAGCTGGAGTTAGAGAAAGCAAAAGAGCGCTCAGAGCGAGTTAGGA 2219  
 Db 2161 GAGAGGAAGAGCTGGAGTTAGAGAAAGCAAAAGAGCGCTCAGAGCGAGTTAGGA 2220  
 Qy 2220 AAGGACAAAGCAATGGCTGGAGCATGTGACAGAGAGCAGCAGCGCCCGGAAACC 2279  
 Db 2221 AAGGACAAAGCAATGGCTGGAGCATGTGACAGAGAGCAGCAGCGCCCGGAAACC 2280  
 Qy 2280 CCAAGAGAGAGCAGACTGAAGAGGAGACAGTGTCAAGAAAGAGAGCGGAAAGAG 2339  
 Db 2281 CCAAGAGAGAGCAGACTGAAGAGGAGACAGTGTCAAGAAAGAGAGCGGAAAGAG 2340  
 Qy 2340 AGCAAGAGCGGAATGTCAAGAGCAGAGTGGGCTTTTCATCCGATCAGAGCAGC 2399  
 Db 2341 AGCAAGAGCGGAATGTCAAGAGCAGAGTGGGCTTTTCATCCGATCAGAGCAGC 2400  
 Qy 2400 TAAAGTGGCCAGCCAGGACCTGTGTCTACACAGAGAAAGCGCGCTTACATTTCTGC 2459  
 Db 2401 TAAAGTGGCCAGCCAGGACCTGTGTCTACACAGAGAAAGCGCGCTTACATTTCTGC 2460  
 Qy 2460 ACAGAGAGGTAAAGTGTATTTACGAGAGGCTGTACCCCTTTGATCCAGAGTGA 2519  
 Db 2461 ACAGAGAGGTAAAGTGTATTTACGAGAGGCTGTACCCCTTTGATCCAGAGTGA 2520  
 Qy 2520 CGATGAGATCAGATCCAGCAGAGATATATGATCATGTGTGATGAAGCGCAGACTGAGA 2579  
 Db 2521 CGATGAGATCAGATCCAGCAGAGATATATGATCATGTGTGATGAAGCGCAGACTGAGA 2580  
 Qy 2580 GCCAGAGTGGCTTGGAGAGAGCTGAAGAGGAGACGGAGTGGTTCCTTGCAAACTATGC 2639  
 Db 2581 GCCAGAGTGGCTTGGAGAGAGCTGAAGAGGAGACGGAGTGGTTCCTTGCAAACTATGC 2640  
 Qy 2640 AGAAGATTCAGAAATAGAGTTCCTCCAGCCCAACAGAGTGCAGTGTGACATC 2699  
 Db 2641 AGAAGATTCAGAAATAGAGTTCCTCCAGCCCAACAGAGTGTGACATC 2700  
 Qy 2700 TGGCCCTGCCCCCAACTGGCTGTGAGAGCCCTGTCTCTTTCAGAGTACCTCTTC 2759  
 Db 2701 TGGCCCTGCCCCCAACTGGCTGTGAGAGCCCTGTCTCTTTCAGAGTACCTCTTC 2760  
 Qy 2760 TGAAGCTTCCACAACCCCAAACTGGGAGACTTCCAGTCCAGTGGCCGACAGCTC 2819  
 Db 2761 TGAAGCTTCCACAACCCCAAACTGGGAGACTTCCAGTCCAGTGGCCGACAGCTC 2820  
 Qy 2820 AAAAGAGAGCCAGAAACGAGCAACTGGAGTACGTGGGCGGCTCAGCTTCTGTACCGT 2879  
 Db 2821 AAAAGAGAGCCAGAAACGAGCAACTGGAGTACGTGGGCGGCTCAGCTTCTGTACCGT 2880  
 Qy 2880 ACCTAGTGTGGCCAGTTACGGCAGAGATCAGCTTTTACCCGACCAAGCCACTGGCTC 2939  
 Db 2881 ACCTAGTGTGGCCAGTTACGGCAGAGATCAGCTTTTACCCGACCAAGCCACTGGCTC 2940  
 Qy 2940 CTGCCCCATCTCCGCTCTGGGCGCAGGCTGAAGAGTGAAGGGCTCAAGCGCAAGCGCT 2999  
 Db 2941 CTGCCCCATCTCCGCTCTGGGCGCAGGCTGAAGAGTGAAGGGCTCAAGCGCAAGCGCT 3000  
 Qy 3000 GTATCCCTGAGAGCCCAAAAGCAACCACTTAATTTTAAACAAAAGTGAAGCTCATCAG 3059  
 Db 3001 GTATCCCTGAGAGCCCAAAAGCAACCACTTAATTTTAAACAAAAGTGAAGCTCATCAG 3060  
 Qy 3060 CGTTCTTGAAACAGCAACATGTGTGTGTTTGAGAAAGTTCAAGAGTCAAGAGGTTGTT 3119  
 Db 3061 CGTTCTTGAAACAGCAACATGTGTGTGTTTGAGAAAGTTCAAGAGTCAAGAGGTTGTT 3120  
 Qy 3120 CCCAAGTCTTACGTGAATCATTTTCAAGGCGCGTAAGAGAAATCCACAAGCATCATAC 3179  
 Db 3121 CCCAAGTCTTACGTGAATCATTTTCAAGGCGCGCGTAAGAGAAATCCACAAGCATCATAC 3180  
 Qy 3180 TGGCCTTACTGAAGGTCTGTAGTCTTAAAGAGAGTGGCTTCCCGCGCCCAAGCCAGC 3239

Db 3181 TGGCCTTACTGAAGGTCTGTAGTCTTAAAGAGTGGCTTCCCGGCGCCAGGCCAGC 3240  
 Qy 3240 CATTCCCGAGAGAGTTTATTGGCATGTACACATACAGAGTTCAGAGCAAGAGATTT 3299  
 Db 3241 CATTCCCGAGAGAGTTTATTGGCATGTACACATACAGAGTTCAGAGCAAGAGATTT 3300  
 Qy 3300 AACCTTTCAGCAAGGAGATGTATTGTGTTTACCAGAAAGATGTGATCTGTGGACGGG 3359  
 Db 3301 AACCTTTCAGCAAGGAGATGTATTGTGTTTACCAGAAAGATGTGATCTGTGGACGGG 3360  
 Qy 3360 AACGTTGGGAGCAAGTCCGAGTCTTCCCTTCACTATGTAGAGCTTAAAGATTGAGA 3419  
 Db 3361 AACGTTGGGAGCAAGTCCGAGTCTTCCCTTCACTATGTAGAGCTTAAAGATTGAGA 3420  
 Qy 3420 GGGCTCTGGAATCTGTGGGAAACAGGAGTTTAGAAAAAAACCTGAATTCGCCAGGT 3479  
 Db 3421 GGGCTCTGGAATCTGTGGGAAACAGGAGTTTAGAAAAAAACCTGAATTCGCCAGGT 3480  
 Qy 3480 TATTGCTTCTTACGCTGTACTGTCTCCGAAACAACTACCTGTGCTTGGGAGAGTGTAT 3539  
 Db 3481 TATTGCTTCTTACGCTGTACTGTCTCCGAAACAACTACCTGTGCTTGGGAGAGTGTAT 3540  
 Qy 3540 TCTGATCCGAAAAAAGAACCCAGGTGATGTGGAGAGAGAACTGCAAGCTCGAGGAA 3599  
 Db 3541 TCTGATCCGAAAAAAGAACCCAGGTGATGTGGAGAGAGAACTGCAAGCTCGAGGAA 3600  
 Qy 3600 AAAGCGCATAGAGGTGTGTTTCCAGCAATATATGTCAAACTTCTAAAGCCCGGACACAG 3659  
 Db 3601 AAAGCGCATAGAGGTGTGTTTCCAGCAATATATGTCAAACTTCTAAAGCCCGGACACAG 3660  
 Qy 3660 CAAAATACCCCACTAGACTACCCAGACCGGAGTGCAGCAGAGTGTGCCAGGTAT 3719  
 Db 3661 CAAAATACCCCACTAGACTACCCAGACCGGAGTGCAGCAGAGTGTGCCAGGTAT 3720  
 Qy 3720 CGGAGTGTACGATTACACCGCCAGAAAGATGACGAACCTAGCTTCAGCAAGGCCAGAT 3779  
 Db 3721 CGGAGTGTACGATTACACCGCCAGAAAGATGACGAACCTAGCTTCAGCAAGGCCAGAT 3780  
 Qy 3780 CATCAAGTCTCTCAACAGAGAGACCCGAGCTGTGTGAAGAGAGAACTCATGTGGCAAGT 3839  
 Db 3781 CATCAAGTCTCTCAACAGAGAGACCCGAGCTGTGTGAAGAGAGAACTCATGTGGCAAGT 3840  
 Qy 3840 TGGGCTTCCATTCATTTATGTAAGCTGACACAGACATGAGACCCAGCCAGCAATG 3899  
 Db 3841 TGGGCTTCCATTCATTTATGTAAGCTGACACAGACATGAGACCCAGCCAGCAATG 3900

RESULT 3  
 AA23909  
 ID AA23909 standard; cDNA, 3723 BP.  
 XX  
 AC AA23909;  
 XX  
 DT 28-FEB-2000 (first entry)  
 XX  
 DE Mouse Eae1 coding sequence.  
 XX  
 KW Mouse; murine; Eae1; Eae2; endocytosis; vesicular trafficking;  
 regulation; actin cytoskeleton; detection; cancer; infection;  
 EH-domain and SH3-domain regulator of endocytosis; anticancer;  
 antiproliferative; antiviral; ss.  
 KW  
 OS Mus sp.  
 XX  
 PN W09955728-A2.  
 XX  
 PD 04-NOV-1999.  
 XX  
 PF 27-APR-1999; 99MO-CA00375.  
 XX  
 PR 27-APR-1998; 98CA-2230201.  
 PR 05-FEB-1999; 99US-0118739.  
 XX

PA (HSCR-) HSC RES & DEV LP.  
 XX Egan SE, Wang W, Sengar A;  
 XX MPI: 2000-052802/04.  
 DR P-PSDB; AAY57444.

XX New nucleic acid encoding Esei and 2 proteins, involved in regulation  
 PT of endocytosis, used e.g. for treating cancer or preventing viral  
 PT infection

XX Claim 6: Page 40-42; 9pp; English.

XX The present sequence encodes mouse Esei. The present invention  
 CC specifically describes mammalian Esei and 2 proteins (I) and their splice  
 CC variants (Ise = EH-domain and SH3-domain regulator of endocytosis). (I)  
 CC are involved in regulation of clathrin-mediated endocytosis (as a complex  
 CC with Eps15 protein), vesicular trafficking and actin cytoskeleton.  
 CC Generally (I) (or its (anti)agonists, mimetics, fragments and inactive  
 CC mutants); (I) -specific antibodies (Ab); sequences antisense to the (I)  
 CC polynucleotide; agents that downregulate expression of Esei genes or  
 CC antagonists of an Esei binding partner are used to treat diseases  
 CC associated with undesirable endocytosis and resulting changes in cellular  
 CC function. Particularly overexpression of Esei is used to block  
 CC clathrin-mediated endocytosis in vivo or in cell cultures, while  
 CC administration of (I) is used to promote endocytosis of selected cells.  
 CC (Anti)agonists of (I) or Ab are used to suppress abnormal proliferation of  
 CC cells that can be stimulated to proliferate by a growth factor receptor;  
 CC and similar compounds (also inactive Esei mutants) can be used to prevent  
 CC viral infection. Endocytosis may also be regulated, in vivo or in cell  
 CC cultures, by forming an Esei-Eps15 complex, then binding dynamin to the  
 CC complex. Generally conditions that can be treated include cancer;  
 CC abnormal cell division or migration; viral infection; or abnormal  
 CC receptor signalling, tissue development or synaptic transmission.

XX Sequence 3723 BP; 1065 A; 941 C; 1025 G; 692 T; 0 other;

XX Query Match 73.2%; Score 3723; DB 21; Length 3723;

XX Best Local Similarity 100.0%; Pred. No. 0;  
 XX Matches 3723; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 259 ATGGCTGATGTTCCACACCTTTGGTGTAGGCTGTGAGCTGGCCATACTGTGGAG 318  
 Db 1 ATGGCTGATGTTCCACACCTTTGGTGTAGGCTGTGAGCTGTGGCCATACTGTGGAG 60  
 QY 319 GAAAGGGCCAGATGACCGCAGTCTTACGCTGAAGCCGATACGGGATTTATTACT 378  
 Db 61 GAAAGGGCCAGATGACCGCAGTCTTACGCTGAAGCCGATACGGGATTTATTACT 120  
 QY 379 GGTGATCAACGAGGAACTTTTTCATCTGGTTACTCAGCCTGTCTTACGACAA 438  
 Db 121 GGTGATCAACGAGGAACTTTTTCATCTGGTTACTCAGCCTGTCTTACGACAA 180  
 QY 439 ATATGGCGCTTACGACATGAATAAGATGAAGATGATCAAGTGAATTTTCATTA 498  
 Db 181 ATATGGCGCTTACGACATGAATAAGATGAAGATGATCAAGTGAATTTTCATTA 240  
 QY 499 GCCATGAAGTTATCAAACTGAAGCTTCAAGATATAGTCTCCCTTCACACTTCCCT 558  
 Db 241 GCCATGAAGTTATCAAACTGAAGCTTCAAGATATAGTCTCCCTTCACACTTCCCT 300  
 QY 559 GTCATGAAACAGCAACAGAGGCTATTTTCAGTGCACACACATTTTGGTATAGAGGATTT 618  
 Db 301 GTCATGAAACAGCAACAGAGGCTATTTTCAGTGCACACACATTTTGGTATAGAGGATTT 360  
 QY 619 GCTAGCATGCGACCACTACAGCTGTGCTCTGTGCGAATGAGCTTCATTCAGTTGTT 678  
 Db 361 GCTAGCATGCGACCACTACAGCTGTGCTCTGTGCGAATGAGCTTCATTCAGTTGTT 420  
 QY 679 GGAATGCTCCACCTTATGATCTTCTGCTTCCCTTCAGACACAGTCTCCCTGGCTAAC 738  
 Db 421 GGAATGCTCCACCTTATGATCTTCTGCTTCCCTTCAGACACAGTCTCCCTGGCTAAC 480

QY 739 GGGGCTCTCCCTGATACAGACCTCTGCTGCTTTGCGATCTCTGACGACCATGGCCA 798  
 Db 481 GGGGCTCTCCCTGATACAGACCTCTGCTGCTTTGCGATCTCTGACGACCATGGCCA 540  
 QY 799 AAGAGTCTCTCCCTGACAGATCGTGGTCCAGGGGCAAAATTAACAATAAGTTACAGAG 858  
 Db 541 AAGAGTCTCTCCCTGACAGATCTGGTCCAGGGGCAAAATTAACAATAAGTTACAGAG 600  
 QY 859 GCACAAATCATTCGATGTCGCGAGCGCCCTTCACAGACAGAAATGGCTGTGCTCACTCA 918  
 Db 601 GCACAAATCATTCGATGTCGCGAGCGCCCTTCACAGACAGAAATGGCTGTGCTCACTCA 660  
 QY 919 TCAAGCTGAAATACAGCAGATTAATTAACAGCCACGACAAACTATAGATGACACTTA 978  
 Db 661 TCAAGCTGAAATACAGCAGATTAATTAACAGCCACGACAAACTATAGATGACACTTA 720  
 QY 979 ACAGTCCCCGAGGAAGAACTATTCATGCAATCAAGTTTACCCGAGCTCAGCTGGCT 1038  
 Db 721 ACAGTCCCCGAGGAAGAACTATTCATGCAATCAAGTTTACCCGAGCTCAGCTGGCT 780  
 QY 1039 TCAATATGAAATCTTTCTGACATTCAGATGCAAGATGAAACTCAGCTGACAGAAATTTATC 1098  
 Db 781 TCAATATGAAATCTTTCTGACATTCAGATGCAAGATGAAACTCAGCTGACAGAAATTTATC 840  
 QY 1099 CTAGCTATGACCTTAATGATGTTGTCATGCTGTGTCAGCCACTGCCCGCTCTGCT 1158  
 Db 841 CTAGCTATGACCTTAATGATGTTGTCATGCTGTGTCAGCCACTGCCCGCTCTGCT 900  
 QY 1159 CCAGAAATACCT 1218  
 Db 901 CCAGAAATACCT 960  
 QY 1219 AGCTCTTCTTCTGATACAGAGCTCTGAGAGCGCTGTCAGAGATGAGACACAG 1278  
 Db 961 AGCTCTTCTTCTGATACAGAGCTCTGAGAGCGCTGTCAGAGATGAGACACAG 1020  
 QY 1279 CCAGAGAAAGAACTGCTGTGACATTTGAAGATTAAGAGCGGGGAACTTGCAGCGAGC 1338  
 Db 1021 CCAGAGAAAGAACTGCTGTGACATTTGAAGATTAAGAGCGGGGAACTTGCAGCGAGC 1080  
 QY 1339 AGTGTGAGCTGGAAGAGCGCGCAAGCGCTTGTGAGACGACGCCAAAGACACAGAG 1398  
 Db 1081 AGTGTGAGCTGGAAGAGCGCGCAAGCGCTTGTGAGACGACGCCAAAGACACAGAG 1140  
 QY 1399 CGGTGCTCTGCTGAGAGCGCGCGAGAGAGAGAAAGCGGAGCGCCAGAGACAG 1458  
 Db 1141 CGGTGCTCTGCTGAGAGCGCGCGAGAGAGAGAAAGCGGAGCGCCAGAGACAG 1200  
 QY 1459 GAGGCGAAGCGGAGCTGAGCTGAGAAAGCACTGAGAAAGCGGGAGCTGAGCGG 1518  
 Db 1201 GAGGCGAAGCGGAGCTGAGCTGAGAAAGCACTGAGAAAGCGGGAGCTGAGCGG 1260  
 QY 1519 CAGCGAGAGAGAGAGAGAGAGATCGAGAGCGCGAGCGCCGAAAAACGGGAACTG 1578  
 Db 1261 CAGCGAGAGAGAGAGAGAGAGATCGAGAGCGCGAGCGCCGAAAAACGGGAACTG 1320  
 QY 1579 GAAAGGCAAGCAACTTGAATGGAACGAAACCGGAGACAGAACTCTGAAATCAAGG 1638  
 Db 1321 GAAAGGCAAGCAACTTGAATGGAACGAAACCGGAGACAGAACTCTGAAATCAAGG 1380  
 QY 1639 AACAGAGACAGAGAGGAGCGGTGCTGAAAGCAAGAGAAAGCACTTGGAGTTTGG 1698  
 Db 1381 AACAGAGACAGAGAGGAGCGGTGCTGAAAGCAAGAGAAAGCACTTGGAGTTTGG 1440  
 QY 1699 TTAAGAGCTCTGAAATGACAAAGATAGCTTAAGAGAAACTTTCAGAGATAGAGTGT 1758  
 Db 1441 TTAAGAGCTCTGAAATGACAAAGATAGCTTAAGAGAAACTTTCAGAGATAGAGTGT 1500  
 QY 1759 CGACTGGCAACCCAGAGGCAAGAAATGAGAGACAGAACTAGAGAGCTTAAGATTT 1818  
 Db 1501 CGACTGGCAACCCAGAGGCAAGAAATGAGAGACAGAACTAGAGAGCTTAAGATTT 1560  
 QY 1819 GCTGAATACCACTTACAGACAGAGTTGAGAGAAATCTAGCAAAATGCTTGAAGACTT 1878

Db 1561 GCTGAATCA CCCACTACAGCAGCAGTTCAGAGAACTTCAGCAAAATGCTTGGAAACCTT 1620  
Qy 1879 ATTCCAGAGAAAACAGATACCTAGTACAGTAAACAAAGTCCAGAGAAACAGTTCAT 1938  
Db 1621 ATTCCAGAGAAAACAGATACCTAGTACAGTAAACAAAGTCCAGAGAAACAGTTCAT 1680  
Qy 1939 AGAGACTGCTTCTTCACTCCCTCAAAAGAGCTTGAAGCAAAAGAGCTGGCCCGCAGCAG 1998  
Db 1681 AGAGACTGCTTCTTCACTCCCTCAAAAGAGCTTGAAGCAAAAGAGCTGGCCCGCAGCAG 1740  
Qy 1999 CTCGGGAGCAGCTGAGCAGAGTGGAGAGAGACAGCTCAAGCTCCAGAGATTGAT 2058  
Db 1741 CTCGGGAGCAGCTGAGCAGAGTGGAGAGAGACAGCTCAAGCTCCAGAGATTGAT 1800  
Qy 2059 GTTTTCAACAACCAAGCTGAAGGAACTGAGAGATACATGCAAAACAGCACTCCGAAG 2118  
Db 1801 GTTTTCAACAACCAAGCTGAAGGAACTGAGAGATACATGCAAAACAGCACTCCGAAG 1860  
Qy 2119 CAGAGTCCCTGAGCAGCAGCAGCTGAGAGCAGAAAGAGAGAGAGAGAGCTGGAG 2178  
Db 1861 CAGAGTCCCTGAGCAGCAGCAGCTGAGAGCAGAAAGAGAGAGAGAGAGCTGGAG 1920  
Qy 2179 TTAG 2238  
Db 1921 TTAG 1980  
Qy 2239 GAGCATGTGAGCAG 2298  
Db 1981 GAGCATGTGAGCAG 2040  
Qy 2299 AAGAGGAG 2358  
Db 2041 AAGAGGAG 2100  
Qy 2359 GACAGAGAGAGTGGGCTTTTCCATCCGATCAGAGAGCAGTAAAGCTGGGCAACCGAGCA 2418  
Db 2101 GACAGAGAGAGTGGGCTTTTCCATCCGATCAGAGAGCAGTAAAGCTGGGCAACCGAGCA 2160  
Qy 2419 CCTGTGCTACACAG 2478  
Db 2161 CCTGTGCTACACAG 2220  
Qy 2479 GTATATTACCGAGAGCTGTACCTTGTGAATCCAGAGTACAGATGATCAACATCCAG 2538  
Db 2221 GTATATTACCGAGAGCTGTACCTTGTGAATCCAGAGTACAGATGATCAACATCCAG 2280  
Qy 2539 CCAGAGATTAAGTCAATGATGATGAAGCAGAGCTGAGAGAGCAGAGATGGTTGAGAG 2598  
Db 2281 CCAGAGATTAAGTCAATGATGATGAAGCAGAGCTGAGAGAGCAGAGATGGTTGAGAG 2340  
Qy 2599 GAGCTGAG 2658  
Db 2341 GAGCTGAG 2400  
Qy 2659 GAGGTTCCCACTCCAGCAGAAACAGTACCGATCTGAGCTGAGCTCCCTGCCCCCAACTG 2718  
Db 2401 GAGGTTCCCACTCCAGCAGAAACAGTACCGATCTGAGCTGAGCTCCCTGCCCCCAACTG 2460  
Qy 2719 GCTTGTGAG 2778  
Db 2461 GCTTGTGAG 2520  
Qy 2779 AACAACTGGGAGAGCTTCAATTCAGTGGGAGCAGAGAGCTCAAAAGAGAGAGAGAGAG 2838  
Db 2521 AACAACTGGGAGAGCTTCAATTCAGTGGGAGCAGAGAGCTCAAAAGAGAGAGAGAGAG 2580  
Qy 2839 GACAACTGGGAGAGCTTCAATTCAGTGGGAGCAGAGCTTCAATTCAGTGGGAGAGAG 2898  
Db 2581 GACAACTGGGAGAGCTTCAATTCAGTGGGAGCAGAGCTTCAATTCAGTGGGAGAGAG 2640  
Qy 2899 CGGAGAGATCAGGCTTTACCCAGCAGAGCAGAGCTGCTTCCCATCTCCGCTCTG 2958  
Db 2641 CGGAGAGATCAGGCTTTACCCAGCAGAGCAGAGCTGCTTCCCATCTCCGCTCTG 2700  
Qy 2959 GGGCAGGGTGAAG 3018  
Db 2701 GGGCAGGGTGAAG 2760  
Qy 3019 AAAGCAACCACTTAAATTTTAAACAAAAGTGAAGCTATCAACCTTCTGAAACAGCAAGAC 3078  
Db 2761 AAAGCAACCACTTAAATTTTAAACAAAAGTGAAGCTATCAACCTTCTGAAACAGCAAGAC 2820  
Qy 3079 ATTGAGTGGTTTGAAG 3138  
Db 2821 ATTGAGTGGTTTGAAG 2880  
Qy 3139 CTCAATTTAGGGCCCGTAAGAGAAATCCACAAAGATCATGCTGAGCTTCTGAAAGTCT 3198  
Db 2881 CTCAATTTAGGGCCCGTAAGAGAAATCCACAAAGATCATGCTGAGCTTCTGAAAGTCT 2940  
Qy 3199 GCTAGCTTAAAG 3258  
Db 2941 GCTAGCTTAAAG 3000  
Qy 3259 ATTGCATGTACATACAG 3318  
Db 3001 ATTGCATGTACATACAG 3060  
Qy 3319 GTGATTTGTGTTACCAAG 3378  
Db 3061 GTGATTTGTGTTACCAAG 3120  
Qy 3379 GAGTCTTCCCTTCTAACTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3438  
Db 3121 GAGTCTTCCCTTCTAACTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3180  
Qy 3439 AAAACAGGAG 3498  
Db 3181 AAAACAGGAG 3240  
Qy 3499 ACTGTTCCCGAGCAACCTACCTTGGCTCTTGGAGAGTGAATCTGATCCGAGAGAGAGAG 3558  
Db 3241 ACTGTTCCCGAGCAACCTACCTTGGCTCTTGGAGAGTGAATCTGATCCGAGAGAGAGAG 3300  
Qy 3559 CCAAGTGTGATGATGAG 3618  
Db 3301 CCAAGTGTGATGATGAG 3360  
Qy 3619 TTTCCAGCAAAATTAATGTCAAACTTTAAGCCCGGAGAGAGAGAGAGAGAGAGAGAG 3678  
Db 3361 TTTCCAGCAAAATTAATGTCAAACTTTAAGCCCGGAGAGAGAGAGAGAGAGAGAGAG 3420  
Qy 3679 CTACCCAG 3738  
Db 3421 CTACCCAG 3480  
Qy 3739 GAG 3798  
Db 3481 GAG 3540  
Qy 3799 GAG 3858  
Db 3541 GAG 3600  
Qy 3859 TATGTAAAGCTGAGCAACAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3918  
Db 3601 TATGTAAAGCTGAGCAACAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660  
Qy 3919 CCCCCTCAGGCTTGAAG 3978  
Db 3661 CCCCCTCAGGCTTGAAG 3720  
Qy 3979 TGA 3981  
Db 3721 TGA 3723

RESULT 4  
AA239025  
ID AA239025 standard; cDNA; 5144 BP.  
XX  
AC AA239025;  
XX  
XX 28-FEB-2000 (first entry)  
DE Mouse Eesll coding sequence.  
XX  
XX Mouse; murine; Eesl; Ees2; endocytosis; vesicular trafficking;  
KW regulation; actin cytoskeleton; detection; cancer; infection;  
KW EH-domain and SH3-domain regulator of endocytosis; anticancer;  
KW antiproliferative; antiviral; ss.  
XX  
OS Mus sp.  
XX  
PN WO955728-A2.  
XX  
XX 04-NOV-1999.  
XX  
XX 27-APR-1999; 99WO-CA00375.  
XX  
XX 27-APR-1998; 98CA-2230201.  
XX  
XX 05-FEB-1999; 99US-0118739.  
XX  
PA (HSCR-) HSC RES & DEV LP.  
XX  
PI Egan SE, Wang W, Sengar A;  
XX  
DR MPI: 2000-052802/04.  
XX  
DR P-PSDB; AAY57449.  
XX  
PT New nucleic acid encoding Eesl and 2 proteins, involved in regulation  
PT of endocytosis, used e.g. for treating cancer or preventing viral  
PT infection -  
XX  
PS Claim 6; Page 59-62; 99pp; English.  
XX  
XX The present invention specifically describes mammalian Eesl and 2  
XX proteins (1) and their splice variants (Ese = EH-domain and SH3-domain  
XX regulator of endocytosis). (1) are involved in regulation of clathrin-  
XX mediated endocytosis (as a complex with Eespl protein), vesicular  
XX trafficking and actin cytoskeleton. Generally (1) (or its (ant)agonists,  
XX mimetics, fragments and inactive mutants); (1)-specific antibodies (Ab);  
XX sequences antisense to the (1) polynucleotide; agents that downregulate  
XX expression of Eesl genes or antagonists of an Eesl binding partner are  
XX used to treat diseases associated with undesirable endocytosis and  
XX resulting changes in cellular function. Particularly overexpression of  
XX Eesl is used to block clathrin-mediated endocytosis in vivo or in cell  
XX cultures, while administration of (1) is used to promote endocytosis of  
XX selected cells. (ant)agonists of (1) or Ab are used to suppress abnormal  
XX proliferation of cells that can be stimulated to proliferate by a growth  
XX factor receptor; and similar compounds (also inactive Eesl mutants) can be  
XX used to prevent viral infection. Endocytosis may also be regulated, in  
XX vivo or in cell cultures, by forming an Eesl-Eespl complex, then binding  
XX dynamin to the complex. Generally conditions that can be treated include  
XX cancer; abnormal cell division or migration; viral infection; or abnormal  
XX receptor signalling, tissue development or synaptic transmission. The  
XX present sequence represents mouse Eesll coding sequence.  
SO Sequence 5144 BP; 1447 A; 1298 C; 1415 G; 964 T; 0 other;

Query Match 71.6%; Score 3639.4; DB 21; Length 5144;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3640; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 259 ATGGCTAGTTCCCAACACTTTCGGTGTAGCTGGATGTCGGGCAATACTGTGGAG 318  
DB 1 ATGGCTAGTTCCCAACACTTTCGGTGTAGCTGGATGTCGGGCAATACTGTGGAG 60

QY 319 GAAAGGCGCAAGCATGACCAAGACTTCCTTAGCCTGAAGCCGATAGCGGATTTATTACT 378  
DB 61 GAAAGGCGCAAGCATGACCAAGACTTCCTTAGCCTGAAGCCGATAGCGGATTTATTACT 120  
QY 379 GGTATCAAGGAGAACTTTTTCCTTCCATCTGGGTTACTCAGCCTGCTTAGACAA 438  
DB 121 GGTATCAAGGAGAACTTTTTCCTTCCATCTGGGTTACTCAGCCTGCTTAGACAA 180  
QY 439 ATATGGCGCTAGCGAGCATGAATTAAGATGAAGATGATCAAGTGAATTTTCCATA 498  
DB 181 ATATGGCGCTAGCGAGCATGAATTAAGATGAAGATGATCAAGTGAATTTTCCATA 240  
QY 499 GCCATGAAGCTTATCAAACTGAAGCTCAAGATATAGCTCCCTCCACACTTCCCT 558  
DB 241 GCCATGAAGCTTATCAAACTGAAGCTCAAGATATAGCTCCCTCCACACTTCCCT 300  
QY 559 GTCATGAAGCAAGCAACAGGCTATTTCCAGTGCACACACTTTGGTATGAGAGGATT 618  
DB 301 GTCATGAAGCAAGCAACAGGCTATTTCCAGTGCACACACTTTGGTATGAGAGGATT 360  
QY 619 GCTAGCATGCCCACTCAGAGCTGTGCTCCTGTCGCAATGGGCTCCATTCCAGTTGTT 678  
DB 361 GCTAGCATGCCCACTCAGAGCTGTGCTCCTGTCGCAATGGGCTCCATTCCAGTTGTT 420  
QY 679 GGAATGTCTCCACCTTAGTATCTTCTGTCCTTCACAGCAGAGTCCCTCCGTGCTAAC 738  
DB 421 GGAATGTCTCCACCTTAGTATCTTCTGTCCTTCACAGCAGAGTCCCTCCGTGCTAAC 480  
QY 739 GGGGCTCTCCCTCATACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 798  
DB 481 GGGGCTCTCCCTCATACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
QY 799 AAGAGTTCTTCCTTCAGCAGATCTGTGTCAGAGGTCACATTTAAACATAAGTTACAGAG 858  
DB 541 AAGAGTTCTTCCTTCAGCAGATCTGTGTCAGAGGTCACATTTAAACATAAGTTACAGAG 600  
QY 859 GCACATCATTCGATGTGCGCCAGGCGCCCTTCACAGCAGAGATGGCTGTGCTCACTCA 918  
DB 601 GCACATCATTCGATGTGCGCCAGGCGCCCTTCACAGCAGAGATGGCTGTGCTCACTCA 660  
QY 919 TCAAGGCTGAATACAGGCTTATTCACAGCAGCAAACTAATGATGAGCACTTA 978  
DB 661 TCAAGGCTGAATACAGGCTTATTCACAGCAGCAAACTAATGATGAGCACTTA 720  
QY 979 ACAGTCCCAAGGAGCAAGCAATCTCATGCAATCAAGTTTACCCAGGCTCAGCTGCT 1038  
DB 721 ACAGTCCCAAGGAGCAAGCAATCTCATGCAATCAAGTTTACCCAGGCTCAGCTGCT 780  
QY 1039 TCAATATGAATCTTCTGACATTCATGAAGTGAAGAACTCACTGCAGAAATTTATC 1098  
DB 781 TCAATATGAATCTTCTGACATTCATGAAGTGAAGAACTCACTGCAGAAATTTATC 840  
QY 1099 CTAGTATGACCTTAATGATGTTGTCATGTCGTGTCAGCAGTCCGCGCTCTGCT 1158  
DB 841 CTAGTATGACCTTAATGATGTTGTCATGTCGTGTCAGCAGTCCGCGCTCTGCT 900  
QY 1159 CCAGAATATCATCT 1218  
DB 901 CCAGAATATCATCT 960  
QY 1219 AGCTTCTTCTGTGTGATCAGAGGCTGCTGAGAGCCGTGTGTCAGAGATGACAGCAG 1278  
DB 961 AGCTTCTTCTGTGTGATCAGAGGCTGCTGAGAGCCGTGTGTCAGAGATGACAGCAG 1020  
QY 1279 CCAGAGAAAGAACTGCTGTGATCAATTTGAAGATGAAGAGCGGGAATCTTGAGCGGAGC 1338  
DB 1021 CCAGAGAAAGAACTGCTGTGATCAATTTGAAGATGAAGAGCGGGAATCTTGAGCGGAGC 1080  
QY 1339 AGTGTGAGCTGAGAAAGCGCGCAAGCGCTCTTGTGAGAGCAGCGCAAGAGCAGAG 1398  
DB 1081 AGTGTGAGCTGAGAAAGCGCGCAAGCGCTCTTGTGAGAGCAGCGCAAGAGCAGAG 1140  
QY 1399 CGGTTTGCTCAGCTGAGAGCGCGCGAGCAGAGAGAAAGCGGAGCGCCAGAGAGCAG 1458



QY	3619	TTTCACCAATATATGTCAACTTCTAAGCCCGGAAACAAGAAATACCCCAATGAG	3678
Db	3361	TTTCCAGCAATTTATGTGCAAACTTCTAAGCCCGGAAACAAGAAATACCCCAATGAG	3420
QY	3679	CTACCCAAAGACCGGATGTCAGCCAGCAGATGTGCGAGTATGGGATGTACGATTAAC	3728
Db	3421	CTACCCAAAGACCGGATGTCAGCCAGCAGATGTGCGAGTATGGGATGTACGATTAAC	3480
QY	3739	GCCAGAACGATGTGCGAACTAAGCTTTCAGAAAGCCAGATCATCAAGTCTTCAACAG	3798
Db	3481	GCCAGAACGATGTGCGAACTAAGCTTTCAGAAAGCCAGATCATCAAGTCTTCAACAG	3540
QY	3799	GAGGACCCGGACTGTGTGGAAGAGAAAGTCAGTGGGCAAGTTGGGCTTTTCCATCAAT	3858
Db	3541	GAGGACCCGGACTGTGTGGAAGAGAAAGTCAGTGGGCAAGTTGGGCTTTTCCATCAAT	3600
QY	3859	TATGTAAAGCTGACCAACAGACATGGAAACCCAGCAGCAATG	3899
Db	3601	TATGTAAAGCTGACCAACAGACATGGAAACCCAGCAGCAATG	3641

RESULT 5  
AAZ34571  
ID AAZ34571 standard; cDNA; 5458 BP.

KM SHD1A gene; human; Down's syndrome; leukemia; cancer;  
KM megakaryocytic abnormality; myeloproliferative disorder;  
KM platelet disorder; neural disorder; thrombocytopenia;  
KM hamatopoietic disorder; cognitive dysfunction; microcephaly;  
KM lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy  
KM 85.

Key	Location/Qualifiers
CDS	267..3929
FT	/*tag= a
FT	

PN W09953062-A2

PD 21-OCT-1999

PF 16-APR-1999; 99WO-US08371.

PR 16-APR-1998; 98US-0082007.

PA (CEDA-) CEDARS SINAI HEALTH SYSTEM  
-----

PI Korenberg JR, Chen X,

DR WPL; 1999-633829/54

XX  
XX

P1 the diagnosis and treatment of myeloproliferative disorders and

Claim 2: Eid 8: 99mn: English

CC This is the nucleotide sequence of full-length cDNA (clone 21)  
CC corresponding to a novel human SH3 gene, termed the SH3DA gene.  
CC that contributes to the development of platelets and the  
CC pathogenesis of leukaemias, both in general and in particular those  
CC involving the megakaryocytic lineage. The SH3DA gene maps to the  
CC small candidate region for low platelets on chromosome 21.  
CC Sequencing of 5 different sites of cDNA clone from foetal brain

CC (see AA234570-74) suggests that at least 3 isoforms exist. The  
CC invention provides methods for the diagnosis and treatment of  
CC megakaryocytic abnormality, myeloproliferative disorder, platelet  
CC disorder, acute leukaemia, neural disorders, thrombocytopenia,  
CC platelet disorder on chromosome 21, low platelets in deletion for  
CC 21, association of gains in chromosome 21 with leukemias, neural  
CC abnormalities, dysfunctions and disorders including brain  
CC malformations and corresponding cognitive dysfunctions,  
CC microcephaly, lissencephaly, and colpocephaly. Methods are also  
CC provided for: suppressing cells unable to regulate themselves;  
CC screening for a somatic alteration in the SH3D1A gene; monitoring  
CC the progress and adequacy of a treatment; monitoring tumour risk  
CC progress or megakaryocytic abnormality, myeloproliferative disorder,  
CC haematopoietic disorder, platelet disorder or leukaemia; and  
CC treatment of a subject (including a prenatal subject) having  
CC megakaryocytic abnormality, myeloproliferative disorder,  
CC platelet disorder, leukaemia or neural disorder using a  
CC nucleic acid that expresses SH3D1A or its antisense nucleic acid.  
XX

Sequence 5458 BP, 1671 A, 1201 C, 1307 G, 1279 T, 0 other;

SQ Sequence 5458 BP; 1671 A; 1201 C; 1307 G; 1279 T; 0 other;

Query Match	69.4%;	Score 3529.4;	DB 20;	Length 5458;
Best Local Similarity	83.7%;	Pred. No. 0;		
Matches 4278;	Conservative	0;	Mismatches 766;	Indels 69;
			Gaps	22;

OY	7	GAGGAGGAGTGGACCGGCGCGGAGGGCGCGACGCTGGTGTGCTCCTAGTACCGCGGCT	66
Db	25	GAGAAAGTGGAGGCGCCGAGGGAGGGAGCGGTACCTGTGTTCTCCGTACTACGCGGCT	84
OY	67	CGCAAGGGAGCATCCCGAGCGGGCTCCGGGACCGGCGAGGACGAGCGGGCGGGCG	126
Db	85	CGCGAGGAAGATCCCGAGCGGGCTCCGGGACG-----GACGAGAGGGCGGGCG	133
OY	127	GGAGTGTGTGCGCGGCTCGGGACTCGGCGTTCTCGC-GGGCGTGGGGCTGTCACTGA	188
Db	134	GGGATGGGTGTGCGGGGCTGCGGCTCTGTGCTCTCCAGCGGGCGGTGAGCGGCACTGA	193
OY	186	TTTGTGTGAGGGGGCGCGCGGCGGACCGCGCGGAGATGAGGGGTGATCAGCAAGGTGA	245
Db	194	TTTGTCCCTGGGGGGGCGCGCGGCGCGCGCGGAGATGAGGGGTGATGCAAGGTAA	251
OY	246	ACGTATATGAAACATNGGCTAGTTTCCACACCTTTGGTGTAGCTGATGTCGTGGG	305
Db	254	AAGTACAGAAACATGGCTAGTTTCCAAACCTTTGGTGGGACGCTGGATATCTGGGC	313
OY	306	CATAACTGTGAGGAGAAAGGCGCAACATGACACAGATTCCTTAGCCTGAACCGATAGC	365
Db	314	CATAACTGTAGGAGAAAGGAGGAGCATGATCAGCAGTTCATAGTTTAAAGCAATATC	372
OY	356	GGGATTTATTACTGGTGTATCAAGCGAGAACTTTTTTCCATCTGGGTTACTGACC	422
Db	374	TGGATTCAATTCTGTGTATCAAGCTTAAGAACTTTTTTTTCAATCTGGGTACTCAAC	433
OY	426	TGTCCTACCAAAATATGGGCGCTAGGGGACATGAATATGATGGAGATGATCAAGT	488
Db	434	TGTTTTACCAAGATATGGGACATGATGAATGAATATGATGAAGATGATCAAGT	493
OY	486	GGAAATTTTCCATAGCGATGAGAGCTTTATCAAACTGAAGCTACAAAGATATCAGTCCCTC	545
Db	494	GGAGTTTTTCCATAGCTATGAAACTTATCAAACTGAAGCTACAAAGATATCAGTCACTC	555
OY	546	CACAATTCCTCCCTGTGCATGAACGCAACAGTGGCTATTTCCAGTGCACCAGCATTTGG	605
Db	554	TGCATTCCTCCCTGTGCATGAACGCAACAGTGGCTATTTCTAGCGACACAGCATTTGG	613
OY	606	TATAGAGAGGATTCGTAGCATGCCACACTCAGACGCTGTGCTCTGTGCCAATGGGCTC	666
Db	614	TATGGAGAGTATCGCAGCATGCCACCGCTTACGCTGTGCTCTCAAGTGCCTCAAGGGATC	673
OY	666	CATTCACATTTGTGGAAATGTCACACCTTAGTATCTTCTGCTCCCTCAGACAGTGGCC	725
Db	674	CATTCACATTTGTGGAAATGTCACACCTTAGTATCTTCTGCTCCACAGACAGTGTGCC	733









Oy	5059	CCGGAAGAAATT	5071	
Db	5104	TCTCAATTAAAT	5116	
RESULT 6				
AA234572				
AA234572		standard; cDNA; 5195 BP.		
AC	AA234572;			
XX				
XX	01-FEB-2000	(first entry)		
DT				
XX				
DE	Human SH3D1A cDNA clone 11.			
XX				
KW	SH3D1A gene; human; Down's syndrome; leukaemia; cancer;			
KW	megakaryocytic abnormality; myeloproliferative disorder;			
KW	platelet disorder; neural disorder; thrombocytopenia;			
KW	haematopoietic disorder; cognitive dysfunction; microcephaly;			
KW	lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy;			
KW	ss.			
OS	Homo sapiens.			
XX				
FH	Key	Location/Qualifiers		
FT	CDS	239..3886		
FT		/*tag= a		
XX				
PN	MO9953062-A2.			
XX				
PD	21-OCT-1999.			
XX				
PF	16-APR-1999;	99WO-U08371.		
XX				
PR	16-APR-1998;	98US-0082007.		
XX				
PA	(CEDA-) CEDARS SINAI HEALTH SYSTEM.			
XX				
PI	Korenberg JR, Chen X;			
XX				
DR	WPI; 1999-633829/54.			
XX	P-PSDB; AAY32156.			
PT	Nucleic acid from the human SH3D1A gene and its products, useful for			
XX	the diagnosis and treatment of myeloproliferative disorders and			
XX	leukaemia			
XX				
PS	Claim 2; Fig 10; 99pp; English.			
XX				
CC	This is the nucleotide sequence of full-length cDNA (clone 11)			
CC	corresponding to a novel human SH3 gene, termed the SH3D1A gene,			
CC	that contributes to the development of platelets and the			
CC	pathogenesis of leukaemias, both in general and in particular those			
CC	involving the megakaryocytic lineage. The SH3D1A gene maps to the			
CC	small candidate region for low platelets on chromosome 21.			
CC	Sequencing of 5 different sizes of cDNA clone from foetal brain			
CC	(see AA234570-74) suggests that at least 3 isoforms exist. The			
CC	invention provides methods for the diagnosis and treatment of			
CC	megakaryocytic abnormality, myeloproliferative disorder, platelet			
CC	disorder, acute leukaemia, neural disorders, thrombocytopenia,			
CC	platelet disorder on chromosome 21, low platelets in deletion for			
CC	21, association of gains in chromosome 21 with leukaemias, neural			
CC	abnormalities, dysfunctions and disorders including brain			
CC	malformations and corresponding cognitive dysfunctions,			
CC	microcephaly, lissencephaly, and colpocephaly. Methods are also			
CC	provided for: suppressing cells unable to regulate themselves;			
CC	screening for a somatic alteration in the SH3D1A gene; monitoring			
CC	the progress and adequacy of a treatment; monitoring tumour risk			
CC	progress or megakaryocytic abnormality, myeloproliferative disorder,			
CC	haematopoietic disorder, platelet disorder or leukaemia; and			
CC	treatment of a subject (including a prenatal subject) having			
CC	megakaryocytic abnormality, myeloproliferative disorder,			
CC	platelet disorder, leukaemia or neural disorder using a			

[illegible]





XX Homo sapiens.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 208..3642  
 FT /tag= a  
 XX  
 XX W09953062-A2.  
 XX  
 XX 21-OCT-1999.  
 XX  
 XX 16-APR-1999; 99WO-US08371.  
 XX  
 XX 16-APR-1998; 98US-0082007.  
 XX  
 PA (CEDA-) CEDARS SINAI HEALTH SYSTEM.  
 PI Korenberg JR, Chen X;  
 XX  
 XX MPI; 1999-633829/54.  
 DR P-PSDB; AAY32154.  
 XX  
 PT Nucleic acid from the human SH3D1A gene and its products, useful for  
 PT the diagnosis and treatment of myeloproliferative disorders and  
 PT leukaemia  
 PS  
 PS Claim 2: Fig 5; 99pp; English.  
 CC This is the nucleotide sequence of full-length cDNA corresponding  
 CC to a novel human SH3 gene, termed the SH3D1A gene, that contributes  
 CC to the development of platelets and the pathogenesis of Leukaemias,  
 CC both in general and in particular those involving the  
 CC megakaryocytic lineage. The SH3D1A gene maps to the small  
 CC candidate region for low platelets on chromosome 21. Sequencing  
 CC of 5 different sizes of cDNA clone (see AAZ34570-74) suggests that  
 CC at least 3 isoforms exist. The invention provides methods for the  
 CC diagnosis and treatment of megakaryocytic abnormality,  
 CC myeloproliferative disorder, platelet disorder, acute leukaemia,  
 CC neural disorders, thrombocytopenia, platelet disorder on  
 CC chromosome 21, low platelets in deletion for 21, association of  
 CC gains in chromosome 21 with leukaemias, neural abnormalities,  
 CC dysfunction and disorders including brain malformations and  
 CC corresponding cognitive dysfunctions, microcephaly, lissencephaly,  
 CC and colpocephaly. Methods are also provided for: suppressing  
 CC cells unable to regulate themselves; screening for a somatic  
 CC alteration in the SH3D1A gene; monitoring the progress and  
 CC adequacy of a treatment; monitoring tumour risk progress or  
 CC megakaryocytic abnormality, myeloproliferative disorder,  
 CC haematopoietic disorder, platelet disorder or leukaemia; and  
 CC treatment of a subject (including a prenatal subject) having  
 CC megakaryocytic abnormality, myeloproliferative disorder,  
 CC platelet disorder, leukaemia or neural disorder using a  
 CC nucleic acid that expresses SH3D1A or its antisense nucleic acid.  
 XX  
 XX Sequence 5199 BP; 1615 A; 1152 C; 1214 G; 1218 T; 0 other;  
 SQ  
 Query Match 60.6%; Score 3080.2; DB 20; Length 5199;  
 Best Local Similarity 79.8%; Pred. No. 0;  
 Matches 4030; Conservative 0; Mismatches 753; Indels 268; Gaps 23;

Db 184 TAGCAGGTAAAGTAAACAGAACCATGCTCAGTTTCAACACCTTTGGAGGAGCTG 243  
 Oy 295 GATGCTGGGCGCATTAAGTGTGAGAGAAAGGCGCAAGCATGACAGCATGCTTGGCTG 354  
 Db 244 GATATCTGGGCGCATTAAGTGTGAGAGAAAGGCGCAAGCATGACAGCATGCTTGGCTG 303  
 Oy 355 AAGCGATAGCGGATTTTATTACTGTGTATCAACGAGGAACTTTTTTCCAACTGGG 414  
 Db 304 AAGCGATATCTGATTCATTAAGTGTGATGATCAAGCTGAAACTTTTTTTCATCTGGG 363  
 Oy 415 TTACTCAGCCTGTCTTACGCAAAATAGGCGGTACGCGAATGATTAAGATGAGAG 474  
 Db 364 TTACTCAGCCTGTCTTACGCAAAATAGGCGGTACGCGAATGATTAAGATGAGAG 423  
 Oy 475 ATGATCAAGTGAATTTTTCATAGCATGAGTATCAAACTGAAGTACAGATAT 534  
 Db 424 ATGATCAAGTGAATTTTTCATAGCATGAGTATCAAACTGAAGTACAGATAT 483  
 Oy 535 CAGCTCCCTCCACAGCTTCCCTGTATGAAACAGCAACCAAGTGTATTTCCAGTGA 594  
 Db 484 CAGCTACCTCTGCACTTCCCTGTATGAAACAGCAACCAAGTGTATTTCCAGTGA 543  
 Oy 595 CCAGATTTGGTATAGAGAGGATTTGATGCAATGCGACACAGCTGTTGCTCCGTG 654  
 Db 544 CCAGATTTGGTATAGAGAGGATTTGATGCAATGCGACACAGCTGTTGCTCCGTG 603  
 Oy 655 CCAATGGGCTCATTTCCAGTTGTAATGTCTCACCTTAGTATCTGTCTCTCA 714  
 Db 604 CCAATGGGCTCATTTCCAGTTGTAATGTCTCACCTTAGTATCTGTCTCTCA 663  
 Oy 715 GCAGCAGTGTCTCCCTGCTGCTTAACGCGGCTCTCTCCATACAGCTTGTGCTGTT 774  
 Db 664 GCAGTGTGCTCCCTGCTGCTTAACGCGGCTCTCTCCATACAGCTTGTGCTGTT 723  
 Oy 775 GCGATCTCTGAGCCACATGCGCAAAAGTCTCTTACGAGATGCGTCCAGGATCA 834  
 Db 724 GCTATCTCTGAGCCACATGCGCAAAAGTCTCTTACGAGATGCGTCCAGGATCA 783  
 Oy 835 CAATTAACACTAAGTTTACAGAGGACCAATCATTTGATGCGAGCGCTCCAGCA 894  
 Db 784 CAATTAACACTAAGTTTACAGAGGACCAATCATTTGATGCGAGCGCTCCAGCA 843  
 Oy 895 GCAGATAGGCTGTGCTCTCAATCAAGGCTGAATTAACGCGCTTATTCACAGCAC 954  
 Db 844 GCAGATAGGCTGTGCTCTCAATCAAGGCTGAATTAACGCGCTTATTCACAGCAT 903  
 Oy 955 GACCAAACTATGAGTGAACACTTAACAGGCTCCAGGCAAACTATTCATGCAATCA 1014  
 Db 904 GACCAAACTATGAGTGAACACTTAACAGGCTCCAGGCAAACTATTCATGCAATCA 963  
 Oy 1015 AGTTTACCAGGCTCAGCTGCTTCAATATGAAATCTTTTGAATGATCAAGATGA 1074  
 Db 964 AGTTTACCAGGCTCAGCTGCTTCAATATGAAATCTTTTGAATGATCAAGATGA 1023  
 Oy 1075 AAATCTACTGAGAGAAATTAATCTTACTATGCACTTAATGATGTTGCTGCTGCT 1134  
 Db 1024 AAATCTACTGAGAGAAATTAATCTTACTATGCACTTAATGATGTTGCTGCTGCT 1083  
 Oy 1135 CAGCCATGCGGCTCTGCTCTCAATTAATCCCTCTCTTCAAGAGGTTGCG 1194  
 Db 1084 CAGCCATGCGGCTCTGCTCTCAATTAATCCCTCTCTTCAAGAGGTTGCG 1143  
 Oy 1195 TCAGCAGTGGATGCTCCGTATTAAGCTCTTCTGTGATGAGAGGCTCTGAGAG 1254  
 Db 1144 TCAGCAGTGGATGCTCCGTATTAAGCTCTTCTGTGATGAGAGGCTCTGAGAG 1203  
 Oy 1255 CCGTGTGAGAGATGAGAGAGCAGC---CAGAGAAAGAACTGCTGTGACATTTGAAGAT 1311  
 Db 1204 CCGTGTGAGAGATGAGAGAGCAGC---CAGAGAAAGAACTGCTGTGACATTTGAAGAT 1263  
 Oy 1312 AAGAACCGGAGAAATTTGAGAGAGCAGTGTGAGAGGCTGAGAGAGCGCGCAGAGCTC 1371





QY 3529 GGGCAGCTATCTGATCCGGAAAAAAGAACCCAGGTGATGTGGAGAGAGACTGCAA 3588  
 Db 3271 GGTACGCTGATTTTGAATCCGAAAAAAGAACCCAGGTGATGTGGAGAGAGACTGCAA 3330  
 QY 3589 GCTTCGAGGAAAAAGCCGAGATAGGCTGTTCCAGCAAAATATATGTCAAACTTTCAAGC 3648  
 Db 3331 GCAGGTGGGAAAAAGCCGAGATAGGCTGTTCCAGCAAAATATATGTCAAACTTTCAAGC 3390  
 QY 3649 CCGGGAACAAGAAAAATCAACCCCACTGACCTACCCAGACCCGAGACCCGAGACCCGAGAC 3708  
 Db 3391 CTTGGACGACGAAAAATCACTCCACAGAGCCACTTAAGTCAACAGACATTTAGGCGACGTG 3450  
 QY 3709 TGGCAGGTGATCGGAGATGATACGATTAACCCGCCAGAACGATGACAACTAGCTTCACG 3768  
 Db 3451 TGGCAGGTGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3510  
 QY 3769 AAGGCGCAATTCATCAACGCTCTCAACAGAGAGACCCGAGACCTGATGAGAAAGAGAGATC 3828  
 Db 3511 AAGGCGCAATTCATCAACGCTCTCAACAGAGAGACCCGAGACCTGATGAGAAAGAGATC 3570  
 QY 3829 AGTGGCAAGTTGAGGCTCTTCCCATCCAAATTAATGTAAGCTGACACAGACATGACCC 3888  
 Db 3571 AATGAGCAAGTGGGCGCTCTTCCCATCCAAATTAATGTAAGCTGACACAGACATGACCC 3630  
 QY 3889 AGCCAGCAATGAATCATATATGTTGATCCATCCCTCAGGCTTGAAGTCTCAAAAGAG 3948  
 Db 3631 AGCCAGCAATGAATCATATATGTTGATCCATCCCTCAGGCTTGAAGTCTCAAAAGAG 3690  
 QY 3949 CCCACTATCCCATATCTCTGCGCCAGAGAGATGATGAGAGATGAGAGCTTGAATGATGAC 4008  
 Db 3691 CCCACTATCCCATATCTCTGCGCCAGAGAGATGATGAGAGATGAGAGCTTGAATGATGAC 3750  
 QY 4009 TTGCAGCATGATCACTACTGCTCTTGGAGTAAAGAACTCACTGACGAGACATTTACCT 4068  
 Db 3751 TTGCAGCATGATCACTACTGCTCTTGGAGTAAAGAACTCACTGACGAGACATTTACCT 3810  
 QY 4069 CATTTGACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4128  
 Db 3811 CATTTGACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3870  
 QY 4129 --AAATTGCAAGATGACAGAGGCTGCTGCTTTGGGGCTTTTCTTATCTCATCAGAC 4186  
 Db 3871 AAAAATATCAAAAAACACAGAGGATGAGGCTGCTTTGGGGCTTTTCTTATCTCATC 3930  
 QY 4187 TGAC--CGGCCCCGCTTCAACAGGCGCTTTCAATAGTTTAAAGATTTTAAATATG 4245  
 Db 3931 TGACTTTTCCCCCCTTGGACAGGCTCTTCAATAGTTTAAATATTTTAAATATAT 3990  
 QY 4246 TATTTAGCTTTTAAATAAATCTCAATCAATTTCTTTGGCTATTTTGGTTTACA 4305  
 Db 3991 TATTTAGCTTTTAAATAAATCAAAATAAATGAATGACTTTTGGCTATTTTGGTTTACA 4050  
 QY 4306 AAAACACCCACTATCAAGAGTGCCTGCTGCGAGACATTTAAATGCTGTTCCGGCGTA 4365  
 Db 4051 AAAAGACCCACTATCAAGAGATG-----CTGCATGCTCTTAAATATTTTCCAAATGT 4105  
 QY 4366 CCGTAAACGAGAGCTGCTGCTTACCTTTGCGGCTTTGCTGAGTGTCCCAACCAATTTG-T 4424  
 Db 4106 CCAATTAATCTGAGACTGATGATGATTTTTCATTTTGTCCAGTGTTCACCAATTAATGCT 4165  
 QY 4425 GTAGTTTGGGGCTGTTCC-----TGCCGTAGAGCAGAGAGATGAGTGTGACTGTTT 4480  
 Db 4166 GCAGTTTGGGGCTTTCCTTCCCTTACATAGAAAGTGCAGAGAGTTCAGTATCTCTGTTT 4225  
 QY 4481 GAAATGTTATGATGAGACTGAGCTGACTATGAGAGGGTTATGCTGCTGAGACATC 4540  
 Db 4226 AAGACGTTATGATGAGAGCCCAATTTAA--GGAAGTGTATGCTGTTTGTGTATC 4284  
 QY 4541 ACGTGTACC--TGTCGCGATGATACATCTGACGAGAGAAATGACT--CTTCCTCATGCG 4598  
 Db 4285 AGCTGTACTTGTGTGAGATGATATACCTCTGTACATAGAAATTTAGTCTTCCATCG 4344  
 QY 4599 TAAACCAACACCGTGTACAGTGTCTCATC-TACTGCATTCATTTTACTTT--GCACAG 4655

Db 4345 CAAACCTTATCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4404  
 QY 4656 TGACCTTGTAGCCACCTGAGAGACAC--CCATGTTTCCGTTTGGCTTCAGATGACCTAG 4714  
 Db 4405 TGACCTTGTAGCCACCTGAGAGACACCTGCTGTTTGTGGCTTCAGATTTATCTGG 4464  
 QY 4715 TTGTCGCCGCTGTTTGTGTTTATTTTCAATCTGAGTGTCTTCAACCATTAACCTAGTA 4774  
 Db 4465 TTGAGTTGCTGTTTGTGTTTGTGGGTTTAAATTTTTCGTTTGCATAGCATTAATAGT 4524  
 QY 4775 AGACCCCACTGCGCCAGCGGTTTACATCATCATCCACCGTC-----TTAGCTCTG 4829  
 Db 4525 AGACCAAC--ACCATGAGGCTGTTAGATCAACGATTCACAGTCTCTTTAGTCTCTG 4583  
 QY 4830 TTACGCTGAAG--TTTATTCAGATGCTTTTATGGA-----TATCTGAAACCAATTAAT 4883  
 Db 4584 TTACATGAAGTTTATTTCAATTTACTTTTCAATGAGAGACCTATTTTGAACCAATTAAT 4643  
 QY 4884 TCTTGCAAGAAAGATGATGAGATGCTCCCTGCAATTAATTTCCAGTGTTCATTT 4943  
 Db 4644 TCTTGCAAGAAAGATGATGAGATGCTCCCTGCAATTAATTT--CCAATGTTTACATTT 4702  
 QY 4944 TTTACTGACTGTGGGCTTGTCTACATTAATGAAATGAGGCTCCGCTGCTGCTG 5003  
 Db 4703 TTTACTGAGACTGTGAAATTTCTACATTAATGAAATGAGGCTCCGCTGCTGCTG 4762  
 QY 5004 TGTGTTAACTGTGCTGTAGCTGAAGCCGTGT-----CTTATGATTTAGTTGGAAGTGC 5060  
 Db 4763 TGTGTT--AGATATGCTGTAGCTGAAGCCGTGTGTTGTTGTTTAAACATGATGGAAGTGC 4821  
 QY 5061 GGAGAGAAATT 5071  
 Db 4822 TCAATAAAAAT 4832

RESULT 8  
 AAS84763  
 ID AAS84763 standard; cDNA; 7435 BP.  
 XX  
 AC AAS84763;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE DNA encoding novel human diagnostic protein #20567.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 OS Homo sapiens.  
 XX  
 PN W0200175067-A2.  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 PI Drmanac RT, Liu C, Tang YT;  
 DR  
 DR MPI: 2001-639362/73.  
 DR P-PSDB: ABG20576.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 PS Claim 1; SEQ ID No 20567; 103bp; English.



XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations in  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. A864197-A894564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [http://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).  
CC  
XX

Sequence 7435 BP; 2190 A; 1690 C; 1829 G; 1726 T; 0 other;

Query Match 51.9%; Score 2638; DB 23; Length 7435;

Best Local Similarity 83.1%; Pred. No. 0;

Matches 3331; Conservative 0; Mismatches 615; Indels 64; Gaps 26;

QY 8 AGGAGAGTGGAGCGGCGGAGGAGGCGGAGCTTGTGCTCCGTAGTACGCGGGCTC 67  
DB 22 AGAAGGTGGATCGCGAGGAGGAGGAGCGTACTTGTGCTCCGTAGTACGCGGGCTC 81  
QY 68 GCAAGGAGATCCCGAGCGGGCTCCGGAGCGCGGAGGAGGAGCGCGGGCGG 127  
DB 82 GCGAGGAGGATCCCGAGCGGGCTCCGGAGCGCGGAGGAGGAGCGCGGGCGG 130  
QY 128 GATGTGTGTCGCGGCTCGGAGCTCGGCTTCTCTGC - GCGGCGTCCGGCTTCACTGAT 186  
DB 131 GATGTGTGTCGCGGCTCGGAGCTCGGCTTCTCTGC - GCGGCGTCCGGCTTCACTGAT 190  
QY 187 TTGTGTGAGGCGCGCGCGGAGCGCGGCGGAGTGAAGCGCTGATCAGCAAGGTAA 246  
DB 191 TTGTGTGAGGCGCGCGCGGAGCGCGGCGGAGTGAAGCGCTGATCAGCAAGGTAA 250  
QY 247 CGTAATGAAGCAATGCTGCTGCTTCCACACCTTGGGTGATGCTGATGCTGAGGCT 306  
DB 251 AGTAACAGAACATGCTGCTGCTTCCACACCTTGGGTGATGCTGATGCTGAGGCT 310  
QY 307 ATAACTGTGAGGAGGAGGCGGAGGAGTGAAGCGAGTTCCTTACCTGAGCGGATAGCG 366  
DB 311 ATAACTGTGAGGAGGAGGAGGAGGAGTGAAGCGAGTTCCTTACCTGAGCGGATAGCG 370  
QY 367 GATTTATTAATGCTGATCAAGGAGGAACTTTTTCCTTCAATCTGCTTACTCAGCTT 426  
DB 371 GATTTATTAATGCTGATCAAGGAGGAACTTTTTCCTTCAATCTGCTTACTCAGCTT 430  
QY 427 GTCTTGACAAATATGCGGCTAGCGGAGCATGATTAAGATGAAGATGATCAAGTG 486  
DB 431 GTCTTGACAAATATGCGGCTAGCGGAGCATGATTAAGATGAAGATGATCAAGTG 490  
QY 487 GAATTTTCCATAGCCATGAAGCTTATCAACTGAAGCTACAGATATAGCTCCCTTCC 546  
DB 491 GAATTTTCCATAGCCATGAAGCTTATCAACTGAAGCTACAGATATAGCTCCCTTCC 550  
QY 547 ACACTTCCCTTGTATGAGAAAGAGCAACAGTGGCTATTTCCAGTCCACAGATTTGGT 606  
DB 551 GCACTTCCCTTGTATGAGAAAGAGCAACAGTGGCTATTTCCAGTCCACAGATTTGGT 610  
QY 607 ATAGAGGAGTGTGCTAGCATGCCACACTCAAGCTGTGCTCTGTCCTGCAATGGGCTCC 666  
DB 611 ATAGAGGAGTGTGCTAGCATGCCACACTCAAGCTGTGCTCTGTCCTGCAATGGGATCC 670

QY 667 ATT - CAGTGTGGAATGTCCTCCACCTTAGTATCTTCTGCTCCACAGAGAGTCC 725  
DB 671 ATTCCAGTGTGGAATGTCCTCCACCTTAGTATCTTCTGCTCCACAGAGAGTCC 730  
QY 726 TCCCTGCTAGCGGGGCTCCTCCGTCATACAGCCCTGCGGCTTGGTGGATCCTGC 785  
DB 731 CCCCCTGCTAGCGGGGCTCCTCCGTCATACAGCCCTGCGGCTTGGTGGATCCTGC 790  
QY 786 AGCCACATGCGCAAGAGTCTTCTTCAAGATGCTGATCAGAGGTCAATTAACAC 845  
DB 791 AGCCACATGCGCAAGAGTCTTCTTCAAGATGCTGATCAGAGGTCAATTAACAC 850  
QY 846 TAAGTTCAGAGGAGCAATATTCATGCTGCGCAGGCGCCCTCC - AGCAGCAATGGG 904  
DB 851 TAAGTTCAGAGGAGCAATATTCATGCTGCGCAGGCGCCCTCC - AGCAGCAATGGG 910  
QY 905 CTGTGCTCAGTCAATCAAGCTGAAATACA - GCGAGTATTCACAGCCACGACAAAA 961  
DB 911 CTGTGCTCAGTCAATCAAGCTGAAATACA - GCGAGTATTCACAGCCACGACAAAA 970  
QY 962 C - TATGATGAGACCTT - AACAGTCCCGAGCAAGAACTTCTCATGATCAATCAAT 1017  
DB 971 CTAATGATGAGACCTTCAATCAAGCTCCCAAGCAAGAACTTCTCATGATCAATCAAT 1030  
QY 1018 TTACCCAGGCTCAGTGGCTCAATATGATGATCTTCTGACATGATCAAGTGAAG 1077  
DB 1031 TTACCCAGGCTCAGTGGCTCAATATGATGATCTTCTGACATGATCAAGTGAAG 1090  
QY 1078 CTCACTGAGAGAAATTTATCTAGCTATGACCTATGATGATGATGATGATGATGAT 1137  
DB 1091 CTCACTGAGAGAAATTTATCTAGCTATGACCTATGATGATGATGATGATGATGAT 1150  
QY 1138 CCACTGCGCGCGGCTCGCTCCAGAAATACATCCCTCCCTTCCAGAGAGTCCCTCC 1197  
DB 1151 CCACTGCGCGCGGCTCGCTCCAGAAATACATCCCTCCCTTCCAGAGAGTCCCTCC 1210  
QY 1198 GCGAGTGGATGTCCTCATAGCTCTTCTTGTGTGATCAGAGGCTGCTGAGAGCG 1257  
DB 1211 GCGAGTGGATGTCCTCATAGCTCTTCTTGTGTGATCAGAGGCTGCTGAGAGCG 1270  
QY 1258 TCGTCAAGATGAGCGAGC - CAGAGAAATCTGCTGTGATTTGAAGTAA 1314  
DB 1271 GTTTTGAAGATGAGCAACAATTTGAAGAAATTTACCTTGAAGTAA 1330  
QY 1315 AAGCGGAGAACTTCAGCGAGGAGTGTGAGTGGAGAGCGCGCAAGGCTCTTGG 1374  
DB 1331 AAGCGGAGAACTTTGAAGCTGCGACCTGGAATGGAAGAAAGCAAGCTCTCTG 1390  
QY 1375 GAGCAGAGCGCAAGAGCAGAGCGGTTGCTCACTGAGCGCGCGCAGAGAGAG 1434  
DB 1391 GAGCAGAGCGCAAGAGCAGAGCGGTTGCTCACTGAGCGCGCGCAGAGAGAG 1450  
QY 1435 AAGAGGCGGAGCGCGCAGAGCGAGGAGCGGAGCTGAGTGGAGAGAGAGCTG 1494  
DB 1451 AAGAGGCGGAGCGCGCAGAGCGAGGAGCGGAGCTGAGTGGAGAGAGAGCTG 1510  
QY 1495 GAGAGAGCGGAGCTGAGCGCGCAGAGCGAGAGAGAGAGAGAGATCGAGAG 1554  
DB 1511 GAGAGAGCGGAGCTGAGCGCGCAGAGCGAGAGAGAGAGAGAGATCGAGAG 1570  
QY 1555 CGCAGGCGCGCAAGCGGAGCT - GGAAGAGCGAGCAACTTGAATGGAGAGAGAG 1613  
DB 1571 CGCAGGCGCGCAAGCGGAGCT - GGAAGAGCGAGCAACTTGAATGGAGAGAGAG 1630  
QY 1614 GAGAGAGAACTCTGATCAGAGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1673  
DB 1631 GAGAGAGAACTCTGATCAGAGAGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1690  
QY 1674 AAGAGAGAGAGCTGAGAGTTTGAAGTCTGATGACAGAGAGAGAGAGAGAGAG 1733  
DB 1691 AAGAGAGAGAGCTTGAAGTTTGAAGTCTGATGAGAGAGAGAGAGAGAGAGAG 1750  
QY 1734 AAGAGAGAGAGAGTATCAGAGTGTGAGCTGAGAGAGAGAGAGAGAGAGAGAG 1793

Db	1751	AGGGAACCTTCAMGATATTCAGATGTGCATTGACCAACCCAAAGCAAGAAATTGAGACAC	1810
Oy	1794	GAACAACTTAGAGAGCTAAGAAATTGCTG--AAATCACCATTACAGACAGCTT--GC	1849
Db	1811	AAACAAATCTTAGAGAGTTGGAATTTCCGGAATATCACCCCAATCTACAGCATTAATTACA	1870
Oy	1850	AGGAATCTCAGCAAA--TGCTTGAAGACTTATTTCCAGAGAAACAGATATCTCAGTACC	1906
Db	1871	TGGAAATCTCAGCAAAATGCTTTGGAAGAACTTAATTCAGAAAACAGATACCTCAATACC	1930
Oy	1907	AGTTAAACAAAGTCCAGCAAAAGTTTCCATAGAGACTCGTCTTTACCTCTCAAAAG-A	1965
Db	1931	AATTAACAAAGTTCCAGCAAAAGTTTCCACAGAGATTCACCTGTTCACCTTAATAAGAA	1990
Oy	1966	GCTTTGAACCAAAAGAGCTGCGCCCGCAGACGCTCCGAGACACTGGAAGAGTGGAG	2025
Db	1991	GCCTTAGAGCAAAAGAACTAGTCCGAGACACTACAGAACCACTGGATGAAGTGGAG	2050
Oy	2026	AGAGAGACCAAGTCAAAAGCTGAGAGATTTGATGTTTCAACAAACAGCTGAAGAACTG	2085
Db	2051	AAAGAACTAGATCAAACTACAGAGATTTGATTTTCAATAACACTGAAGAACTA	2110
Oy	2086	AGAGAGTACATAGCAAAACGCACTTCCAGAAACAGAGTCCCTTGAAGCAGCCGACTG	2145
Db	2111	AGAAATAATCACATATAGCAAACTCCAGAAACAAAGTCCATGGAAGCTGAACGACTG	2170
Oy	2146	AAGCAGAAAGCAGCAGAGAGAAAGAGCTGAGATTGAGAAACAAAGAAAGAGCGCTCAG	2205
Db	2171	AAACAGAAAGAACAGAACCAAAAGATCTATGATTTGAAAAACAAAAAGAAAGAACCA	2230
Oy	2206	AGACGAGTTACAGAAAGAGCAACAGCAATGGCTGGAGCATGTGCAGCAGAG--GAGCAG	2262
Db	2231	AGACGAGCTCAGAAAGAGGACAAGAGTGGCTTGAACATGTGCACAGAGAGCAGCAT	2290
Oy	2263	CCACGCCCCCGGAAACCCCAACAGAGAGACAGACTTGAACAAGGAAAGACAGTGTCAAGAG	2332
Db	2291	CAGAGCCCAAGAAAACTCCAAGAAAGAAAACTGAAGAGAGAGAAAGTGTCAAAAAAG	2350
Oy	2323	AAGGA-GCCGGAAACAGAGAGCCAAAGCCGGAATGTCAACAGCAGAGTCCGCTTTTCCA	2381
Db	2351	AAGGATGCGAGAGAAAAAGGCAAACAGAGACCAAGACCAAGAGCTGGTCCGCTTTTCCA	2410
Oy	2382	TCCGCATCAGGAGCCAGCTAAGCTGGCCACCCAGCAGACCCTGCTTACACAGAGAAAGG	2441
Db	2411	TCAACACCAABAACACACTAAGCAGCTGTCCAGGACCCCTGTCACACAGAAAAAGG	2470
Oy	2442	CCCGCTTACCATTTTCTCACAGAGAGTGTAAAGTGTATATATACCGAGCGCTGTACC	2501
Db	2471	TCCACTTACCATTTTCTGCACAGGAAAAATGAAAGGTGTATATCCGGGCACTGTACC	2530
Oy	2502	CTTTGAATCCGAAGTCAACATAGATCAACATCCAGCCAGAGAAATATAGCAT-----	2555
Db	2531	CTTTGAATCCGAAGCCATATGAAATCACTATCCAGCCAGAGACATATGATATGTTAA	2590
Oy	2556	-----GCTGATGAAAAAGCCAGACTGAGAGCCAGAGTGGCTTGGAGAGACTGAA	2606
Db	2591	AGGGGAATGGGTGTGATGAAGCCAAATGGAAGAACCCGCTGGCTTGGAGAGAAATTAA	2650
Oy	2607	AGGGAAGCGGGATGTTCCCTGCAAACTATGCAAAAAAGATTCAGAAAATGAGTTCC	2666
Db	2651	AGGAAGACAGGGGTGTCTCCCTGCAAACTATGCAAGAAAAATCCAGAAAAATGAGTTCC	2710
Oy	2667	CACCTCAGCCAAACACAGTGAACCATCTGAATCTGCCCTCGCCCCCAACCTGGCTCTCG	2726
Db	2711	CGCTCAGATGAACACAGTGACTGATTCMAACATCTGCCCTCGCCCAACCTGGCTTCCG	2770
Oy	2727	TGAGACCCCTCTCTCTTGGCAGAGACCTCTTCTGAAGCCTCCACAAACCCCAACACTG	2786
Db	2771	TGAGACCCCTCTCTTGGAGAGTAACTCTTCAAGACCTCCACGACCCCTTAATAACTG	2830
Oy	2787	GGCAGACTTCAGTTCCACGTGGCCAGACGCTCAACAGAGAACGAAACGACACTG	2846

Db	2831	GGCGCACTTACGCTCCAGCTGGCGCCACGACGACGAAATGAAACCAAGAAACGATTAAC	2829
Qy	2847	GGATACGTGGGCGGCTCAG - CCTTCTTGAACCGTACCTTAAGCTGGGCCAGTTACGCGAGA	2905
Db	2891	GGATGCACTGGGCGACGAGGCCCTTCTCTCAACCGTTCCAAAGTCCGGCCAGTTAAGGCGAGA	2950
Qy	2906	GATACGCTTTACCCCAAGCCACAGCCACTGGGTCTCTCCCAATCTCCCGTCCCTGGGCCAGG	2965
Db	2951	GGTCGGCTTTACTCCAGCCACGGCCACAGCGCTCTCCCGCTCTCTGTGTAAAGCCAGG	3010
Qy	2966	GTGAAAAGGTGGAAGGCGCTACCAAGCGCAAGCCCTGTATCCCTGAGAGACC - AAAAAGAC	3024
Db	3011	GTGAAAAGGTGAGGGGCTTCANAGCTCANAGCCCTTATCTTCTTGGAGAGCCAAAAAGAC	3070
Qy	3025	AACCACTTAAATTTTAAACAAAGTGAAGTCATCACCGTTCTTGGAAACAGACAGACATGTGG	3084
Db	3071	AACCACTTAAATTTTAAACAAAGTGAATGTCATCACCGTCTTGGAAACAGACAGACATGTGG	3130
Qy	3085	TGGTTTGAAGAGTTCAAGGTCAAGAGGTTGGTTTCCCAAGTCTTACAGTAACTCATTT	3144
Db	3131	TGGTTTGAAGAGTTCAAGGTCAAGAGGTTGGTTTCCCAAGTCTTACAGTAACTCATTT	3190
Qy	3145	TCAGAGCGCCGTAAGGAAATCAACAGCAT - CGATCTGGCCCTACTGTAAGGCTCGCTAG	3203
Db	3191	TCAGAGCGCCATAGAAAGTTTACACATAGGAATTCGTGTTCTTCAGAGAGTCTCTGCTAG	3250
Qy	3204	TCTAAAGAGAGTGGCTTCCCGGCCCGCCCAAGCCAGCATTTCCCGGAGAA - GAGTTTATTG	3262
Db	3251	TCTAAAGAGAGTGGCTTCCCGGCCCGCCCAAGCCAGCATTTCCCGGAGAA - GAGTTTATTG	3310
Qy	3263	CCATGTACACATACAGAGATTCTGAG - CAAGAGATTTAACCTTT - CAGCCMAGGAT	3318
Db	3311	CCATTTACACTTACGAAAGTTCTGAAGGCAAGGAGATTTTAACTTTTCAAGCCAGGAGAT	3370
Qy	3319	GTGATTGGTTTACCAAGAAAGATGTGTGACTGGTGAACGGGAAAC - GGTGGCGACAATC	3377
Db	3371	GTGATTGGTTTACCAAGAAAGATGTGTGACTGGTGAACGAAGAACTAGTGGCGACAGAGC	3420
Qy	3378	CGAGTCTTCCCTTCTAACTATGTGAGGCTTAAAGATTCAAGGGCTCTGGAACGTCTGG	3437
Db	3431	CGAGTCTTCCCTTCTAACTATGTGAGGCTTAAAGATTCAAGGGCTCTGGAACGTCTGG	3490
Qy	3438	GAAAACAGGAGTTTACGAAAAAAACCTGAAATTG - CCCAGTTATTGCTTCTTACGCTG	3486
Db	3491	GAAAACAGGAGTTTACGAAAAAAACCTGAAATTGCCCCAGGTTATTGCTCTATACACCG	3550
Qy	3497	CTACTGTCTCCCGAACACTACCTCGCTCTGGGAGTGAATCTGAATCCGGAAGAAAGA	3556
Db	3551	CCACGCGCCCGACACACTCACTCTCGCCCTGTGTAGCTGATTTTGATCCGAAAAAGAA	3610
Qy	3557	A - CCCAGGTGATGTGTGGAAGAGAACTGCAAGCTCGAGGAAAAAGGCCAGATAAGG	3615
Db	3611	ACCCAGGTGATGTGTGGAAGAGAGCTGCAAGCATGTGGAAAAAGGCCAGATAGGC	3670
Qy	3616	TGGTTTCCAGCAAAATTATGTCAAACTTTCTAAGCCCGGAACAAACAATCAACCCCACT	3675
Db	3671	TGGTTTCCAGCAAAATTATGTCAAACTTTCTAAGCCCGGAACAAACAATCAACCCCACT	3730
Qy	3676	GAGCTACCCCAAGACCGAGATGCAAGCCACAGCATGTGTGCAAGTATGGGATGTAGCATAC	3735
Db	3731	GAGCACTTAAGTCAACAGGATTAAGCGGCAAGTGTGCAAGTATGGGATGTAGCATAC	3790
Qy	3736	ACCGCCCGAAGCATGTAGCAAACTAGCTTACAGAA - AGGCCAGATCATTAACGTCGAA	3794
Db	3791	ACCGCCCGAAGTACAGATGAGCTGGCTTCAACAAAGGCCAGATCATTAAGTCTCAA	3850
Qy	3795	CAAGAGAGACCCGGAAGTGTGTGAAAAGGAGAGTCACTGGGCAAGTTGGCTTTCCCATC	3854
Db	3851	CAAGAGAGACCCGGAAGTGTGTGAAAAGGAGAGTCACTGGGCAAGTTGGCTTTCCCATC	3910
Qy	3855	CAATTATGTAAAGTGAACACAGCATGTGACCCCAAGCCAGCAATGAATATATGTGTGCC	3914
Db	3911	CAATTATGTGAAGTGAACACAGCATGTGAGGAGCCCAAGCCAGCAATGTGTTAGACTTAC	3970



PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249219.
PR	17-NOV-2000;	2000US-0249220.
PR	17-NOV-2000;	2000US-0249221.
PR	17-NOV-2000;	2000US-0249222.
PR	17-NOV-2000;	2000US-0249223.
PR	17-NOV-2000;	2000US-0249224.
PR	17-NOV-2000;	2000US-0249225.
PR	17-NOV-2000;	2000US-0249226.
PR	17-NOV-2000;	2000US-0249227.
PR	17-NOV-2000;	2000US-0249228.
PR	17-NOV-2000;	2000US-0249229.
PR	17-NOV-2000;	2000US-0249230.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256179.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	11-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.
XX	(HUMA-) HUMAN GENOME SCI INC.	
PI	Rosen CA, Barash SC, Ruben SM;	
XX	WPI; 2001-488781/53.	
DR	P-PSDB; AAM43519.	
XX	New isolated nucleic acids and polypeptides, useful for diagnosing,	
PT	treating and/or preventing human diseases and disorders -	
PS	Claim 1; SEQ ID NO 33; 664pp + Sequence Listing; English.	
XX	The invention relates to human polynucleotides (AA163803-AA164012) and	
CC	the encoded proteins (AAM434497-AAM43660) useful for preventing, treating	
CC	or ameliorating medical conditions e.g. by protein or gene therapy. The	
CC	genes were isolated from a range of human tissues disclosed in the	
CC	specification. The nucleic acids, proteins, antibodies and (ant)agonists	
CC	are useful in the diagnosis, treatment and prevention of: (a) cancer,	
CC	e.g. breast and ovarian cancer and other cancers of the adrenal gland,	
CC	bone, bone marrow, breast, gastrointestinal tract, liver, lung, or	
CC	urogenital; (b) immune disorders e.g. Addison's disease, allergies,	
CC	autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,	
CC	Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative	
CC	colitis; (c) cardiovascular disorders such as myocardial ischaemias;	
CC	(d) wound healing; (e) neurological diseases such as viral, bacterial, fungal	
CC	epilepsy; and (f) infectious diseases such as viral, bacterial, fungal	
CC	and parasitic infections.	
CC	Note: The sequence data for this patent did not form part of the	
CC	printed specification, but was obtained in electronic format directly	
CC	from WIPO at ftp.wipo.int/pub/published_pct_sequences.	
XX		
XQ	Sequence 3466 BP; 1056 A; 829 C; 861 G; 710 T; 10 other;	
	Query Match 43.1%; Score 2188.8; DB 22; Length 3466;	
	Best Local Similarity 84.7%; Pred. No. 0;	
	Matches 2525; Conservative 0; Mismatches 427; Indels 30; Gaps 5	
OY	7 GAGGAGAGGTGGACGCCCGCGGAGGCGCGCAGCTTGTTGTCCTCGTAGTACGGCGCTT 66	
Db	35 GAGAGAAGTGGAGGCGCCACAGGAGGAGGAGCGTAGTGTGTTGTCCTCGTAGTACGGCGCTT 94	
OY	67 CGCAAGGAGCATCTCCGAGCGCGGCTCCGGAGAGCGCCGGGAGGCAGGACGGCGCGGCG 126	
Db	95 CGCAGGAAGAATCTCCGAGCGGCTCCGGAGG-----GACGAGAGCGCGGCGC 143	
OY	127 GGAGTGTGTGCGCGGCTCGAGCTCGGCGTTCTTCG-CGCGCGTGGGCGCTCACATTGA 185	
Db	144 GCGATGTGTGTGCGGCGGTGGCGCTCTCGTCTCCCTTCACAGCGCGCGCGGAGCGGACTTGA 203	
OY	186 TTGTGTGAGGGCGCGCGCGCACCTCGCGGAGATGAGGCGTGCATCGACAAGTGA 245	

Db	204	TTTGTCCTCGAGGCGGAGCGGGAACCCGCCGAGATGAGGCGTGCATTGACAAGTAA	263
Qy	246	ACGTAATAGAACCATGGCTCAGTTTCCACACCTTTGGTGTAGCTCGATGTCTGGCC	305
Db	264	AAGTAACAGAACCATGGCTCAGTTTCCACACCTTTGGTGTAGCTCGATGTCTGGCC	323
Qy	306	CATPACTGTGAGAAAGAGGCCAAGCATACGAGAGTTCCCTTACCTCGAAGCCGATAGC	365
Db	324	CATPACTGTGAGAAAGAGGCCAAGCATATGACAGTTTCCATAGTTTAAAGCCAAATATC	383
Qy	366	GGGATTTATTAAGTGTGATCAAGCGAGAACTTTTTTTCATCTGGGTTACTCGACC	425
Db	384	TGAGTTTATTACTGTGATCAAGCTTAGAAACTTTTTCATCTGGGTTACTCGAAC	443
Qy	426	TGCTTTAGCAAAATATGGGCGCTAGCGGACATGAAATACGATGGAAGATGATCAAGT	485
Db	444	TGTTTTAGCAAAATATGGGCACTAGCTGACATGAAATATGTAATGGAATGATCAAGT	503
Qy	486	GGAAATTTTCATAGGCATGAGCTTATCAAACTGAAGTACAGGAATATCGACTCCCTC	545
Db	504	GGAAATTTTCATAGGCATGAGCTTATCAAACTGAAGTACAGGAATATCGACTCCCTC	563
Qy	546	CACACTTCCCCCTGTATGATAAAGCAACAGCAACAGTGGCTATTTCAGTGCACGAGATTGG	605
Db	564	TGCATTTCCCCCTGTATGATAAAGCAACAGCAACAGTGGCTATTTCAGTGCACGAGATTGG	623
Qy	606	TATAGAGGAGATTGTAGATGACACACTCAAGCTGTGCTCTGTGCCAATGGGCTC	665
Db	624	TATGGAGGATGCGCAGCATGCCACCGCTTACAGCTGTGCTCCAGTCCAAATGGGATC	683
Qy	666	CATTCCAGTTGTTGGAAATGTCTTCACCTTAGTATCTTGTCTCTCCAGCAGCAGTGGC	725
Db	684	CATTCCAGTTGTTGGAAATGTCTTCACACCTTAGTATCTTGTCTCCAGCAGCAGTGGC	743
Qy	726	TCCCTGGCTTAACGGGGCTCTCCCGTCAATAGAGCTGCGCTGGTGTGGCATCTGCG	785
Db	744	CCCCCTGGCTTAACGGGGCTCTCCCGTCAATAGAGCTGCGCTGGTGTGGCATCTGCG	803
Qy	786	AGGCACATGGCCAAAGATTTCTCTTCAGCAGATCTGGTCCAGGGTCACAATTAAACAC	845
Db	804	AGGCACATGGCCAAAGATTTCTCTTCAGTATCTGGTCCAGGGTCACAATTAACAC	863
Qy	846	TAAATTCAGAAAGGCAAGCTATTTGATGTGGCAGTGTCCACAGTGGGAGATGGGC	905
Db	864	TAAATTCAGAAAGGCAAGCTATTTGATGTGGCAGTGTCCACAGTGGGAGATGGGC	923
Qy	906	TGTGCTCACTCATCAAGGCTGAATAACGAGAGTTTCAACAGCCAGCAAAACAT	965
Db	924	TGTGCTCACTCATCAAGGCTGAATAACGAGCAATTTCAATATCTCATGCAAAACAT	983
Qy	966	GAGTGGACATTAAAC-----AGGTCCCGCAGCAAGATATTTCATGCAATC	1013
Db	984	GAGTGGACATTAAACAGGTTCCGTTTAGTGTCCCAAGGAATAATTTCTTATGCAATC	1043
Qy	1014	AAGTTTACCCAGGCTCAGCTGGCTTCAATATGAAATCTTTGTGACATTTGATCAAGATG	1073
Db	1044	AAGTTTACCCAGGCTCAGCTGGCTTCAATATGAAATCTTTGTGACATTTGATCAAGATG	1103
Qy	1074	AAAACTACTGCAAGAAATTTATCTTACTATGCACTTAATGATGTTGGCATGCTGG	1133
Db	1104	AAAACTTACAGAGAGAAATTTATCTTGGCAATGCACTCATGATGATGCTGCTGG	1163
Qy	1134	TCAGCCACTGCGCGCGTCTGGCTCAAAATACATCCCTCTTCAAGAGGTTGCG	1193
Db	1164	CCAAACACTGCGCGCTGTCTGTCTCCAGAAATACATTCACCTTTTAAAGAGTTGCG	1223
Qy	1194	CTCCGAGTGGGATGTCCGCTCAATAGCTCTTCTTGTGTGATCAGAGGCTGCTGAGA	1253
Db	1224	ATCTGGCAGTGTATATCTGTCAATAGCTCAACATCTGTAGATCAGAGGCTACAGAGGA	1283
Qy	1254	GCGCTGTCAAGAGATGAGCAGAGC--CAGAGAAAGAACTGCTGTGACATTTGAAGA	1310

Db 1284 ACCAGTTTGAAGATGACACACAACTTGAAGAAATTAAGTCTGTAAGCTTTGAGA 1343  
 Qy 1311 TAAGAAAGCGGAGAACTTCAAGGAGGCACTGTGAGCTGTGAGAACGCCGCCAACCGCT 1370  
 Db 1344 TAAGAAAGCGGAGAACTTGAACGTGTGCAACTGTGAATCTGGAAGAAAGAAAGCAACTCT 1403  
 Qy 1371 CTGGAAGCAGCAGCGCAAGAGCAGGAGCGTTGGCTCAGCTGAGACCGCGCAGAGAGA 1430  
 Db 1404 CTGGAAGCAGCAGCGCAAGAGCAGGAGCGCTGTGCGCCAGCTGTGAGCGGCGGAGAGAGA 1463  
 Qy 1431 GAGAAAGACCGGAGCGCCAGAGAGCAGAGCCAGCGCACCTGAGCTGTGAGAGAGA 1490  
 Db 1464 GAGAAAGACCGTGTGAGCGCCAGAGAGCCAGAAAGCAACTGTGAAGTGAAGAGA 1523  
 Qy 1491 GCTGGAAGACAGCGGAGCGTGTGAGCGGCGCAGAGAGAGAGAGAGAGAGAGATGCA 1550  
 Db 1524 ACTGGAAGAAAGCGGAGCGCTGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1583  
 Qy 1551 GAGGCGGAGGCGGCAAAAGCGGAACTGGAAGCGAGCGCAAACTTGAATGGGAAAGGAA 1610  
 Db 1584 GAGGCGGAGGCGTGTGCAAAAGCGGAACTTGAAGCGCAAACTTGAAGTGGGAAAGGAA 1643  
 Qy 1611 CCGGAGACAGAACTCTGTAATCAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1670  
 Db 1644 TCGAAGCGAAGAACTACTAAATCAAGAAAGCAAGAGAGAGAGAGAGAGAGAGAGAG 1703  
 Qy 1671 GCGAAG 1730  
 Db 1704 AGCAAG 1763  
 Qy 1731 AGAAGAGAACTTCAAGATATCAGGTGTGCACTGGCAAGCCAGAGAGAGAGAGAGAGAG 1790  
 Db 1764 AGAAGGAACTTCAAGATATCAGGTGTGCACTGGCAAGCCAGAGAGAGAGAGAGAGAG 1823  
 Qy 1791 CAGGAG 1850  
 Db 1824 CAGGAG 1883  
 Qy 1851 GGAATCTCAGCAATGTGTTGAGAGACTTATTCAGAGAGAAAGAGAGAGAGAGAGAGAG 1910  
 Db 1884 GGAATCTCAGCAATGTGTTGAGAGACTTATTCAGAGAGAAAGAGAGAGAGAGAGAGAG 1943  
 Qy 1911 AAAACAAGTCCAG 1970  
 Db 1944 AAAACAAGTCCAG 2003  
 Qy 1971 GGAAG 2030  
 Db 2004 AGAAGCAAAAG 2063  
 Qy 2031 GACCAAGTCAAAAGTGTGAGAGAGATGTGTTTCAACCAACAGCTGAAGAGAGAGAGAG 2090  
 Db 2064 AACTAGATCAAAAGTGTGAGAGAGATGTGTTTCAACCAACAGCTGAAGAGAGAGAGAG 2123  
 Qy 2091 GATTCATAGCAAAAG 2150  
 Db 2124 AATTCATAGCAAAAG 2183  
 Qy 2151 GAAAG 2210  
 Db 2184 GAAAG 2243  
 Qy 2211 AGTTTCAAG 2267  
 Db 2244 AGCTCAAG 2303  
 Qy 2268 CCCCCGAG 2327  
 Db 2304 ACCAAG 2363  
 Qy 2328 GCGGAG 2387  
 Db 2364 TGGGAG 2423

Qy 2388 TCAGAGCAGACTAAGCTGCGCAACCCAGAGACCTGTGTTACACAGAGAGAGAGAGAGAG 2447  
 Db 2424 CCAAG 2483  
 Qy 2448 TACCAATTTCTGCAAG 2507  
 Db 2484 TACCAATTTCTGCAAG 2543  
 Qy 2508 ATCCAG 2567  
 Db 2544 ATCCAG 2603  
 Qy 2568 CCAG 2627  
 Db 2604 CCAAG 2663  
 Qy 2628 TGCAAACTATGCAAG 2687  
 Db 2664 TGCAAACTATGCAAG 2723  
 Qy 2688 CGATCTGACATCTGCGCCCTGCGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2747  
 Db 2724 TGATTCAG 2783  
 Qy 2748 AGTGAAGCTTTCTGAG 2807  
 Db 2784 AGTGAAGCTTTCTGAG 2843  
 Qy 2808 GCCAG 2867  
 Db 2844 GCCAG 2903  
 Qy 2868 TTCTCTGAG 2927  
 Db 2904 CTCTCTGAG 2963  
 Qy 2928 AGCAG 2969  
 Db 2964 GGCAG 3005

RESULT 10  
 ABR43498  
 ID ABR43498 standard; cDNA; 3319 BP.  
 XX  
 AC ABR43498;  
 XX  
 DT 05-JUN-2002 (first entry)  
 XX  
 DE DNA encoding novel central nervous system protein #78.  
 XX  
 KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;  
 KW hyperproliferative disorder; neoplasia; cardiovascular disorder;  
 KW cardiac arrest; cerebrovascular disorder; ischemia; angiogenesis;  
 KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;  
 KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;  
 KW adenocarcinoma; reproductive system disorder; testicular feminisation;  
 KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;  
 KW respiratory disorder; renal disorder; kidney failure; blood disorder;  
 KW myocardial infarction; wound healing; cell proliferation; skin aging;  
 KW food additive; food preservative; gene therapy; gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200155318-A2.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US01332.  
 XX  
 PR 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 21-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 25-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254078.  
PR 05-JAN-2001; 2001US-0259678.  
  
(HUMA-) HUMAN GENOME SCI INC.  
PA  
XX  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-581633/65.  
DR P-PSDB; AAU87168.  
XX  
XX  
PT New isolated nucleic acid encoding a protein for diagnosing,  
PT preventing, treating or ameliorating medical conditions and used as  
PT food additives or preservatives -  
XX





QY	1835	TACGACGACGAGTTGACAGGATCTCACCAATGCTTGAAGAATTATTCACAGAAACAGA	1894
Db	1750	TACGACCAACATTACAGGAATCTCAGCAATGCTTGAAGAATTATTCAGAAACAGA	1809
QY	1895	TACTCAGTGAACAGTTTAAAACAAGTCACACAGAACAGTTTGCATAGAGACTCGCTTTA	1954
Db	1810	TACTCAATGACCAATTTAAAACAAGTTCAGAGAACAGTTTGCACAGAGATTCACTTGTTA	1869
QY	1955	CCCTCAAAAAGCCTTTGGAAAGCAAAGAGTGGCCCGGACAGACTCCGGAGACGCTGG	2014
Db	1870	CACTTAAAAAGCCTTTAGAAAGCAAAAAGAACTAGCTTCGACAGCACTACAGAACCAACTGG	1929
QY	2015	ACGAGGTGAGAGAGAGACCAAGGTCAAAAGTCGAGAGATTGATGTTTCAACAAACAGC	2074
Db	1930	ATGAAGTGGAGAAAAGAAACTAGATCAAAAATACAGAGATTGATATTTTCAATTAATCAAC	1989
QY	2075	TGAAGAACTGAGAGAGATACATATGCAAAACAGCAATCCAGAAAGACAGAGTCCCTGAGG	2134
Db	1990	TGAAGAACTTAAAGAAATATCAATATGAACAACTCCAGAAAGCAAAAGTCATAGGAG	2049
QY	2135	CAGCGCACTGAAAGCAGAAAAGACAGAGAGAGAGAGAGCCTGAGATTAGAGAAACAAAAG	2194
Db	2050	CTGAACCACTGAAACAGAAAAGAACAAAGATCATAGAAATTAGAAAACAAAAG	2109
QY	2195	AAGCGCTCAGAGACGAGTTCAAGAAAAGGACAACGAATGGCTGGAGCATGTGCGACAG	2254
Db	2110	AAGAAAGCCAAAGACGAGCTCAAGAAAAGGACAAAGATGGCTGAGACATGTGCGACAG	2169
QY	2255	AG--AGACGACACGCCCCCGGAAACCCACACAGAGAGACAGACTGAAGAGGCAAGACA	2311
Db	2170	AGGACGAGCATCAGAGACCAAGAAACCTCCACGAAGAGAAAACCTGAAGAAAGGAGAG	2229
QY	2312	GTGTCAAGAAAGAGAGCGGAAGAGAGAGCCCAAGCCGGAATTCAGACAAACAGAGTC	2371
Db	2230	GTGTCAAAAAGAGATGCGAGAGAAAAGGCAAAACAGAAACCAACAGCAAGCTGCGTCC	2289
QY	2372	GGCTTTTCCATCCGCATCAGAGACCAAGCTAAGCTGGCCACCCAGGCACCTTGCTTACCA	2431
Db	2290	GGCTTTTCCATCAACACCAAGAACCAAGCTAAGCAAGCTGTCCAGGACCTTGCTCACTG	2349
QY	2432	CAGAGAAAGACCCGCTTACCATTTCTGCAACAGAGAGTGTAAAGTGATATTTACACAG	2491
Db	2350	CAGAAAAGGTCCACTTACCATTTCTGCAAGGAAATGTAAAAAGTGGTATTTACCGG	2409
QY	2492	CGCTGTACCCCTTTGAATCCAGAAAGTCAAGATGATCAACATCCAGCCAGAGATATAG	2551
Db	2410	CACGTGACCCCTTTGAATCCAGAAAGCATGATGAATTCATTCAGCCAGAGACATAG	2469
QY	2552	TCATGTGTGATGAAAGCCACACCTGAGAGCCAGAGATGGCTTGAAGAGACTGAAAGGGA	2611
Db	2470	TCATGTGTGATGAAAGCCAAACTGGAAGAACCCGCGCTGTGGAGAGATTTAAAAAGGA	2529
QY	2612	AGAGGGGATGGTTCCTCGGAAACTATGACAGAAAAGTTCCAGAAAATGAGGTTCCCACTC	2671
Db	2530	AGAGGGGTGGTTCCTCGCAAACTATGACAGAAAATCCAGAAAATGAGGTTCCCGCTC	2589
QY	2672	CAGCAAAACAGTGAACCGATCTGACATCTGCCCCCTCCCAACTGAGCTCTGCGGTGAGA	2731
Db	2590	CAGTGAAAACCGATGATCGATTCAACATCTGCCCCCTCCCAACTGCGCTTGAGAGA	2649
QY	2732	CCCTGCTCTCTTTGACAGTACCTCTTTCTGAGCCCTTCACAAACCCCAACAATCTGGGAG	2791
Db	2650	CCCCCGCCCTTTGGCAGTAACTCTTCAGAGCCCTTCACAGACCCCTTAATTAATCTGGGCG	2709
QY	2792	ACCTTCAGTTCACGTGGCCCAAGAGGCTCAAAGAGAAAGCAGAAACGGAACAATCTGGGATA	2851
Db	2710	ACCTTCAGTTCACGTGGCCCAACAGACCGAATAGAAAACAGAAAGGATTAATCTGGGATG	2769
QY	2852	CGTGGGCGGCTCAGCCTTCTCTGACCGTACCTTATGATCTGGACAGTTACGCGACAGATAG	2911
Db	2770	CATGGGACGCCACACCTCTCTCACCGTTCAAGTGCAGGCTTAAGCAGAGAGTCCG	2829
QY	2912	CCTTTACCCCAAGCCACAGCCATGCGCTCTCTCCCATCTCCCGTCTGCGGACAGGCTGA	2969

D6	2830	CCTTTACTCCAGGCCACGGCGCACTG	GCTTCTCCCGGTCTCCTGTGTAAGGCAGCTGA	2887
<hr/>				
RESULT 11				
ID	AAK94139			
XX	AAK94139	standard; cDNA; 2131 BP.		
AC	AAK94139;			
XX				
DT	06-NOV-2001	(first entry)		
XX				
DE	Human full-length cDNA; SEQ ID NO: 2646.			
XX				
KM	Human; full length cDNA; cDNA synthesis; oligo-capping; ss.			
OS	Homo sapiens.			
PX	EPIJ30094-A2.			
PN				
FD	05-SEP-2001.			
PF				
PR	07-JUL-2000; 2000EP-0114089.			
RR	08-JUL-1999; 99JP-0194486.			
SR	11-JAN-2000; 2000JP-0118774.			
PA	02-MAY-2000; 2000JP-0183765.			
HELI-	(HELI-) HELIX RES INST.			
Ota T,	Nishikawa T., Isogai T., Hayashi K., Ishii S., Kawai Y.;			
Wakamatsu A,	Sugiyama T., Nagai K., Kojima S., Otsuki T., Koga H;			
WI:	2001-524255/58.			
P-PDDB;	AAM93229.			
PT	830 Primers useful for synthesizing full length cDNA clones and their			
use in genetic manipulation -				
Claim 8;	SEQ ID NO 2646; 1380bp + sequence listing; English.			
The invention relates to primers for synthesising full length cDNA clones. 830 cDNA molecules encoding a human protein have been isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have been determined. Primers for synthesising the full length cDNA are useful for clarifying the function of the protein encoded by the cDNA. The full length clones were obtained by construction of full length enriched cDNA libraries that were synthesised by the oligo-capping method. The primers enable the production of the full length cDNA easily without any special methods. The present sequence is a full length human cDNA of the invention.				
Note: The sequence data for this patent did not form part of the printed specification, but was obtained in CD-ROM format directly from EPO.				
Sequence 2131 BP; 694 A; 469 C; 526 G; 442 T; 0 other:				
Query Match	31.7%; Score 1610; DB 22; Length 2131;			
Best Local Similarity	85.5%; Pred.No. 0;			
Matches 1832; Conservative	0; Mismatches 225; Indels 15; Gaps 3;			
QY	45 GTTGCTCCGTAGTAGCGCGGCTTCGGAAGGAGAATCCCAGCGGCTCCGGAGCGGCGG	104		
Db	1 GTTGCTCCGTAGTAGCGCGGCTTCGGAAGGAGAATCCCAGCGGCTCCGGAGCGG---	57		
QY	105 GAGGAGGAGCGCGGCGGCGGGAATGTGTGCGCGGCTCGGAGCTGCGGGCTTCGCG	164		
Db	58 -----CAGAGAGCGCGGCGGCGGATGTGTGCGGCTTCGCGCTCTCTGCTCTCC	109		
QY	165 -GCGGCGTGTGCGGCTGTGATTGTTGTGAGGGGCGGCGCGCACCGCCGAGAT	223		
Db	110 AGCGGCGGTGAGCGGCGCATGTATTGTCTCCGTGGGGCGGACAAGCGGAGAT	169		
QY	224 GAGGCGTGTGATGAGCAGGTGAAGTAAATGAGAACCATGCTCAAGTTTCCACACCTTTGG	283		

Db 170 GAGCGCTGCGATTAGCAAGTAAAGTAACAGAACCATAGGCTCAGATTCCAAACCTTTTG 229  
 QY 284 GTGGTACCTGGAGTGTCTGGGGCCATACTGTGAGGAAAGGGCCAGCATGACAGCAGT 343  
 Db 230 GTGGACGCTGGATATCTGGGCCATTAAGTAGAGAAAGGAGACATGATCAGCAGT 289  
 QY 344 TCCCTTACCTGAAAGCCGATAGCCGGATTTATTAAGTGTGATCAAGCCGAGAACTTTTTT 403  
 Db 290 TCCATAGTTAAAGCCATATCTGGATTCTGATTAATCTGATCAAGCTAAGAACTTTTTT 349  
 QY 404 TCCATCTGGTTTACCTCAGCTGTCTTAGCACAATAATATGGGGCTTACCGGATGAATA 463  
 Db 350 TTCAATCTGGTTTACCTCAACCTGTTTACGACAGATATGGGCACTAGCTGACATGATA 409  
 QY 464 ACGATGAGAGATGATCAAGTGAATTTTTCATAGCCATGAAAGCTTATCAACTGAGC 523  
 Db 410 ATGATGAGAGATGATCAAGTGAATTTTTCATAGCCATGAAAGCTTATCAACTGAGC 469  
 QY 524 TACAGGATATCAGCTCCCTCCACACTTCCCTGTATGAAACAGCAACAGTGGCTA 583  
 Db 470 TACAGGATATCAGCTCCCTCCACACTTCCCTGTATGAAACAGCAACAGTGGCTA 529  
 QY 584 TTTCCAGGACAGCAGCTTTGGTATAGAGGATTTGCTAGATGCCACACTCAGCAGCTG 643  
 Db 530 TTTTCAAGCAGCAGCAGCTTTGGTATAGAGGATTTGCTAGATGCCACACTCAGCAGCTG 589  
 QY 644 TTGCTCTGTGCAATGAGGCTTCATTCAGTGTGGATGTCTCAACCTTAGATCTT 703  
 Db 590 TTGCTCTGTGCAATGAGGCTTCATTCAGTGTGGATGTCTCAACCTTAGATCTT 649  
 QY 704 CTGTCCTCCAGCAGCAGTGGCTCCCTGGCTAACGGGGCTCCCTCCCTATCAGCCTC 763  
 Db 650 CTGTCCTCCAGCAGCAGTGGCTCCCTGGCTAACGGGGCTCCCTCCCTATCAGCCTC 709  
 QY 764 TGCGTGGTTGGCCGATCTGCAAGCCATAGGCAAGAGTTCCTTCCAGAGATCTG 823  
 Db 710 TGCGTGGTTGGCTCCTGATCTGCAAGCCATAGGCAAGAGTTCCTTCCAGAGATCTG 769  
 QY 824 GTCCAGGCTCAATTAACCTAAGTTACAGAAAGCAGCAATCTTGTAGTGGCCAGCG 883  
 Db 770 GTCCAGGCTCAATTAACCTAAGTTACAGAAAGCAGCAATCTTGTAGTGGCCAGCG 829  
 QY 884 CCGCTCAGCAGCAGAAATGGGCTGTGCTCAGTCAAGGCTCAATTAAGCAGAGTAT 943  
 Db 830 TCCACCAAGTGGCAGAGTGGGCTGTGCTCAGTCAAGGCTCAATTAAGCAGAGTAT 889  
 QY 944 TCAACAGCAGCAGCAAACTATGAGTGAACAATTACAGGTTCCCGCAGGCAAGAACTATT 1003  
 Db 890 TCAATAGTCATGACAAAACCTATGAGTGAACAATTACAGGTTCCCGCAGGCAAGAACTATT 949  
 QY 1004 TCATGCAATCAAGTTTACCCAGGCTCAGCTGGCTTCAATTAAGAACTTTTTCAGCAATTG 1063  
 Db 950 TTATGCAATCAAGTTTACCCAGGCTCAGCTGGCTTCAATTAAGAACTTTTTCAGCAATTG 1009  
 QY 1064 ATCAAGATGAAAACTCACTGCAAGAAATTTATCTAGTATGCACTAATGATGTTG 1123  
 Db 1010 ATCAAGATGAAAACTCACTGCAAGAAATTTATCTAGTATGCACTAATGATGTTG 1069  
 QY 1124 CCATGTCTGTGACGCACTGCGCCGCTGCTGCTCCAGAAATACATCCCTTCTTCA 1183  
 Db 1070 CTATGTCTGTGCAACCACTGCACTGCTGCTGCTCCAGAAATACATCCCTTCTTCA 1129  
 QY 1184 GAAAGATTCGCTCCGGGAGTGGGATGCCGTCAATAGTCTCTTCTGTGATCAGAGG 1243  
 Db 1130 GAAAGATTCGCTCCGGGAGTGGGATGCCGTCAATAGTCTCTTCTGTGATCAGAGG 1189  
 QY 1244 TGCTGAGGACCGTGTGTCAGAGATGAGCAGCAGC---CAGAGAGAAACTGCTGTGA 1300  
 Db 1190 TACCAAGAGAACCAATTTTGAAGATGAAACAACAATTTGAAAGAAATTTACTGTAA 1249  
 QY 1301 CATTTGAAGATTAAGAGGGGAGAACTTTCAGGCAAGCAGTGTGAGCTGAGAGAGGCC 1360

Db 1250 CATTGAAGATTAAGAGCGGAGAACTTTGACGTGGCAACTGGAACTGGAGAAAGAA 1309  
 QY 1361 GCCAAGCGCTCTTGGAGCAGCAGCGCAAAAGAGCAGGAGCGGTTGGCTCAGCTGAGCGC 1420  
 Db 1310 GCGAAGCTCTCTGGAGACAGCAGCGCAAGAGCAGAGCGCTGCGCCAGCTGAGCGGG 1369  
 QY 1421 CCGAGCAGAGAGAGAAAGAGCGGAGCGCCAGAGCAGAGAGCCCAAGCGCAGCTGAGC 1480  
 Db 1370 CCGAGCAGAGAGAGAAAGAGCGGAGCGCCAGAGCAGAGCGCCCAAGAGCACTGGAAC 1429  
 QY 1481 TGGAGACAGCTGGAGAGAGCAGCGGAGCTGAGCGGAGCGGAGAGAGAGAGAGAG 1540  
 Db 1430 TGGAGACCACTGGAGAAAGCAGCGGAGCTAGAACCGCAGAGAGAGAGAGAGAGAG 1489  
 QY 1541 AGAGATCGAGAGCGCGAGCGCCGCAAAACGGGAACTGGAAAGCGCAGCACTTGAAT 1600  
 Db 1490 AAGAAATTTGAGAGCGGAGAGCGCTCAAAAACGGGAACTTTGAAAGCAACGCAACTTGA 1549  
 QY 1601 GGGAGCGGAGCCGAGAGCAGAGAACTCCTGAATCAGAGAACAGAGAGCAGGAGCGC 1660  
 Db 1550 GGGAGCGGAGATCGAGAGCAGAGAACTCCTGAATCAGAGAACAGAGAGCAGGAGCA 1609  
 QY 1661 TGGTCTGAGAGCAAGAGAGAACTCTGAGATTGAGATTGAAAGCTCTGATGACAAA 1720  
 Db 1610 TTGTACTGAAAGCAAGAAAGAAAGCTTTGAAATTTGAAATTGAAAGCTCTTAATGATAAA 1669  
 QY 1721 AGCATCAGCTAGAGAGAAACTTACAGATATCAGAGTGTCCATGCGCAACCCAGAGGCA 1780  
 Db 1670 AGCATCAACTGAGAGAGAACTTACAGATATCAGAGTGTCCATGCGCAACCCAGAGGCA 1729  
 QY 1781 AATTTGAGAGCAGCAAGCTAGAGAGCTAAGAAATTTGTAATCAGCAATCACCCTTACAGC 1840  
 Db 1730 AATTTGAGAGCAGCAAGCTAGAGAGCTAAGAAATTTGTAATCAGCAATCACCCTTACAGC 1789  
 QY 1841 AGCAGCTGAGAGATCTCAGCAATGCTTGGAGAACTTATTCAGAGAAACAGATCTCA 1900  
 Db 1790 AACAAATTCAGAGAACTCAGCAAAATGCTTGGAGAACTTATTCAGAGAAACAGATCTCA 1849  
 QY 1901 GTGACAGTTAAACAAATGCTCAGAGAAACAGTTTGGACATGAGAGCTGCTTACCTCA 1960  
 Db 1850 ATGACCAATTAACAAATGCTCAGAGAAACAGTTTGGACATGAGAGCTGCTTACCTCA 1909  
 QY 1961 AAAGAGCTTGGAGAGCAAGAGAGCTGCGCCGAGCAGCACTCCGAGAGCAGCTGAGCAG 2020  
 Db 1910 AAAGAGCTTGGAGAGCAAGAGAGCTGCGCCGAGCAGCACTCCGAGAGCAGCTGAGCAG 1969  
 QY 2021 TGGAGAGAGAGCCAGGTCAAAGCTGCGAGAGATGATGTTTCAACAACCACTGAAAG 2080  
 Db 1970 TGGAGAGAGAGAACTGATCAAAACCTACAGAGATTGATTAATTAATATCAGCTGAAAG 2029  
 QY 2081 AACTGAGAGAGATATGATGCAAAACAGCAACTCCAGAAAGCAGAGGTCCTGAGAGCAGCG 2140  
 Db 2030 AACTGAGAGAGATATGATGCAAAACAGCAACTCCAGAAAGCAGAGGTCCTGAGAGCAGCG 2089  
 QY 2141 GACTGAG 2182  
 Db 2090 GACTGAG 2131

## RESULT 12

AA234573 standard; cDNA; 2079 BP.

AA234573;

01-FEB-2000 (first entry)

Human SH3D1A cDNA clone 5.

DE SH3D1A gene; human; Down's syndrome; leukemia; cancer;  
 XX megakaryocytic abnormality; myeloproliferative disorder;  
 KW platelet disorder; neural disorder; thrombocytopenia;  
 KW haematopoietic disorder; cognitive dysfunction; microcephaly;

lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy;  
 ss.  
 Homo sapiens.  
 Key Location/Qualifiers  
 CDS 136..2079  
 FT /tag= a  
 MO9953062-A2.  
 21-OCT-1999.  
 16-APR-1999; 99WO-US08371.  
 16-APR-1998; 98US-0082007.  
 (CEDA-) CEDARS SINAI HEALTH SYSTEM.  
 Korenberg JR, Chen X;  
 MPI, 1999-633829/54.  
 P-PSDB; AA32157.  
 Nucleic acid from the human SH3D1A gene and its products, useful for  
 the diagnosis and treatment of myeloproliferative disorders and  
 leukemia  
 Claim 2: Fig 12; 99pp; English.  
 This is the nucleotide sequence of a cDNA clone, termed clone 5,  
 corresponding to a novel human SH3 gene, termed the SH3D1A gene,  
 that contributes to the development of platelets and the  
 pathogenesis of leukaemias, both in general and in particular those  
 involving the megakaryocytic lineage. The SH3D1A gene maps to the  
 small candidate region for low platelets on chromosome 21.  
 Sequencing of 5 different sizes of cDNA clone from foetal brain  
 (see AA24570-74) suggests that at least 3 isoforms exist. The  
 invention provides methods for the diagnosis and treatment of  
 megakaryocytic abnormality, myeloproliferative disorder, platelet  
 disorder, acute leukaemia, neural disorders, thrombocytopenia,  
 platelet disorder on chromosome 21, low platelets in deletion for  
 21, association of gains in chromosome 21 with leukaemias, neural  
 abnormalities, dysfunctions and disorders including brain  
 malformations and corresponding cognitive dysfunctions.  
 microcephaly, lissencephaly, and colpocephaly. Methods are also  
 provided for: suppressing cells unable to regulate themselves;  
 screening for a somatic alteration in the SH3D1A gene; monitoring  
 the progress and adequacy of a treatment; monitoring tumour risk  
 progress or megakaryocytic abnormality, myeloproliferative disorder,  
 haematopoietic disorder, platelet disorder or leukaemia; and  
 treatment of a subject (including a prenatal subject) having  
 megakaryocytic abnormality, myeloproliferative disorder,  
 platelet disorder, leukaemia or neural disorder using a  
 nucleic acid that expresses SH3D1A or its antisense nucleic acid.  
 Sequence 2079 BP; 700 A; 451 C; 495 G; 433 T; 0 other;  
 Query Match 31.2%; Score 1584.4; DB 20; Length 2079;  
 Best Local Similarity 85.8%; Pred. No. 0;  
 Matches 1783; Conservative 0; Mismatches 291; Indels 4; Gaps 2;

304 GCCATTAACGTGAGAGAAAGGGCCAGCATGACGAGAGTTCTTACCTGAAAGCCGATA 363  
 DB GCCATTAACGTGAGAGAAAGGGCCAGCATGATGACGAGTTCTTAAAGCCGATA 240  
 364 GCGGATTTATTACTGTGATCAAGCGAGAACTTTTTCACATCTGGGTTACCTCAG 423  
 DB TCTGATTCATTACTGTGATCAAGCGAGAACTTTTTCATCTGGGTTACCTCAA 300  
 424 CCGTCTTAGCACAATAATGAGGCGCTAGCGGACATGATAACATGAGAGATGATCAA 483  
 DB CCGTCTTAGCACAATATGAGGCGCTAGCGGACATGATAACATGAGAGATGATCAA 360  
 484 GTGGAATTTTCATAGCCATGAGCTTATCAACTGAGCTACAGAGATATCACTCC 543  
 DB GTGGAATTTTCATAGCTATGAGCTTATCAACTGAGCTACAGAGATATCACTCC 420  
 544 TCCACACTTCCCTCTGTCATGAAACAGCAACAGTGGCTATTTCCAGTGCACGACTTT 603  
 DB TCCACACTTCCCTCTGTCATGAAACAGCAACAGTGGCTATTTCCAGTGCACGACTTT 480  
 604 GGTATGAGAGGATTTGTAGCATGCGACACTCAAGCTGTGCTCGTGGCCAAAGGCG 663  
 DB GGTATGAGAGGATTTGTAGCATGCGACACTCAAGCTGTGCTCGTGGCCAAAGGCG 540  
 664 TCCATTCCAGTTGTTGGAATGTCTCAACCTTATGATCTTCTCCCTCCAGCAGCAGTG 723  
 DB TCCATTCCAGTTGTTGGAATGTCTCAACCTTATGATCTTCTCCCTCCAGCAGCAGTG 600  
 724 CTTCCCTGGCTTAACGGGGCTCTCCGCTATACAGCTCTGCTGCTTGGCCATCT 783  
 DB CTTCCCTGGCTTAACGGGGCTCTCCGCTATACAGCTCTGCTGCTTGGCCATCT 660  
 784 GGAGCACAATGGCCAAAGATTTCTTCAGCATGCTGCTGAGGTTACATTAAC 843  
 DB GGAGCACAATGGCCAAAGATTTCTTCAGCATGCTGCTGAGGTTACATTAAC 720  
 844 ACTAAGTTACAGAAAGGCAATATCTTCAGATGCTCCAGCGCCCTCCAGCAGCAATGG 903  
 DB ACTAAGTTACAGAAAGGCAATATCTTCAGATGCTCCAGCGCCCTCCAGCAGCAATGG 780  
 904 GCTGTGCTCAGTATCAAGGCTGAATATCAGGAGTTATTCAGCAGCAGCAAACT 963  
 DB GCTGTGCTCAGTATCAAGGCTGAATATCAGGAGTTATTCAGCAGCAGCAAACT 840  
 964 ATGATGACACTTAACAGTCTCCCGCAGCAAGATTTCTCAAGCAATCAAGTTACCC 1023  
 DB ATGATGACACTTAACAGTCTCCCGCAGCAAGATTTCTCAAGCAATCAAGTTACCC 900  
 1024 CAGGCTCAGGCTGCTTCAATATGGAATCTTTCTGACATGATCAAGATGAAATCTCA 1083  
 DB CAGGCTCAGGCTGCTTCAATATGGAATCTTTCTGACATGATCAAGATGAAATCTCA 960  
 1084 GCAGAGAAATTTATCTGATGATGACCTTAATGATGATGATGATGATGATGATGAT 1143  
 DB GCAGAGAAATTTATCTGATGATGACCTTAATGATGATGATGATGATGATGATGATGAT 1020  
 1144 CCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1203  
 DB CCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080  
 1204 GGGATGTCGATATAGCTCTTCTGATGATGATGATGATGATGATGATGATGATGAT 1263  
 DB GGGATGTCGATATAGCTCTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 1140  
 1264 GAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1320  
 DB GAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1200  
 1321 GAGAACTTGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1380  
 DB GAGAACTTGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1260  
 1201 GAGAACTTGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1260



QY 306 CATAACTGGAGAGAAAGGCGCAGCATGACGAGCTTCTTACCTGGAAGCGATAGC 365  
 Db 394 CATTAATCTAGAGAGAAAGCGAAGCATGATGACGAGTTCATAGTTTAAAGCAATATC 453  
 QY 366 GGGATTATTATCTGATGATCAGCGAGAACTTTTTCATCTGGGTACCTCAGCC 425  
 Db 454 TGGATTATTAATGAGTCAAGCTAGAACTTTTTCATCTGGGTACCTCAGCC 513  
 QY 426 TGTCTTAGCAAAATATGGCGCTGACGACATGAATTAAGATGAGATGATCAAGT 485  
 Db 514 TGTATTAGCAAGATATGGCGACTAGCTGACATGAATTAATGATGAGAAATGATCAAGT 573  
 QY 486 GGAATTTTCCATAGAGCATGAAGCTTATCAAACTGAAGCTACAAAGATATCAGCTCCCTC 545  
 Db 574 GGAGTTTTCATAGTATGAACTTATCAAACTGAAGCTACAAAGATATCAGCTACCTCC 633  
 QY 546 CAGACTTCCCTGCTGATGAAACAGCAACCGAGCTATTTCCAGTGCACAGCATTTGG 605  
 Db 634 TGCACATCCCTGCTGATGAAACAGCAACCGAGTGTCTATTTTCAGGCGACAGCATTT-- 691  
 QY 606 TATAGAGAGATGCTAGCATGCGACCACTCAGAGCTGTGCTCCTGTCGAATGGGCTC 665  
 Db 692 ----- 691  
 QY 666 CATTCAGTTTGAATGTCTCCACCTTAGTATCTTCTCCTCAGCAGAGCTGCC 725  
 Db 692 ----- GCAAGTGTGCT 702  
 QY 726 TCCCTGCTTAGAGGGGCTCTCCGCTCATACAGCTCTGCTGCTTGGCGATCTCTGC 785  
 Db 703 CCCCCTGGCTTAAAGGGGCTCTCCCTGTATACAACTCTGCTGCAATTTGCTATCTCTGC 762  
 QY 786 AGCCACATGGCCAAAGAGTCTTCTCTCAGCAGATCTGGCTCCAGGCTCAAAATTAACAC 845  
 Db 763 AGCCACATGGCCAAAGAGTCTTCTCTTATAGATCTGGCTCCAGGCTCAAAATTAACAC 822  
 QY 846 TAAATTACAAAGAGCACAATCATTTGATGTGCGCAGCGCCCTCCAGCAGCAAGATGGCC 905  
 Db 823 TAAATTACAAAGAGCACAATCATTTGATGTGCGCAGGCTCCAGCAGTGGCAGAGTGGCC 882  
 QY 906 TGTGCTCAGTATCAAGGCTGAATATCAGGCGATTTATCAAGCAGCAAGCAAACTAT 965  
 Db 883 TGTCTCAGTATCAAGGCTGAATATCAGGCGATTTATCAAGCAGCAAGCAAACTAT 942  
 QY 966 GAGTGGCACTTAAAGAGTCCCGAGGAGAACTATTTCTGATGATCAAGTTTACCCCA 1025  
 Db 943 GAGTGGCACTTAAAGAGTCCCGAGGAGAACTATTTCTGATGATCAAGTTTACCCCA 1002  
 QY 1026 GGTCTAGCTGCTTCAATATGAAATCTTTCTGACATTTGATCAAGATGAGAACTCAGTCC 1085  
 Db 1003 GGTCTAGCTGCTTCAATATGAAATCTTTCTGACATTTGATCAAGATGAGAACTCAGTCC 1062  
 QY 1086 AGAAGAAATTTATCTAGCTATGACCTAATTTGATTTGCTGATCTGCTGAGCACTGCC 1145  
 Db 1063 AGAAGAAATTTATCTAGCTATGACCTAATTTGATTTGCTGATCTGCTGAGCACTGCC 1122  
 QY 1146 GCGCGTCTGCTCCAGAAATACATCCCTCTCTCAGAAAGTTCGCTCCGCGAGTGG 1205  
 Db 1123 ACCTGCTCTGCTCCAGAAATACATCCCTCTTTTGAAGAGTTTCATCTGCGAGTGG 1182  
 QY 1206 GATGCTCGTATAGCTCTTTCTGTGATCAGAGCTGCTGAGAGAGCGTCTGACAG 1265  
 Db 1183 TATATCTGTATAGCTCAACATCTGTATGATCAGAGCTGCTGAGAGAGCGTCTTTTGA 1242  
 QY 1266 GATGAGCAGCAGC---CAGAGAAAGAACTGCTGTGACATTTGAAGATTAAGAGGGGGA 1322  
 Db 1243 AGATGAACAACAATTAAGAAAGAAATTAAGTGAAGTGAAGTGAAGAGGGGA 1302  
 QY 1323 GAACCTTGAGAGAGAGAGTGTGAGTGAAGAGCGCGCAAGCGCTTTGAGAGCAGCA 1382  
 Db 1303 GAACCTTGAGAGAGAGTGTGAGTGAAGAGCGCGCAAGCGCTTTGAGAGAGCA 1362  
 QY 1383 GCGCAAGAGAGAGAGCGTGTGCTGAGTGAAGCGCGCGAGCAGAGAGAGAAAGAGCG 1442

Db 1363 GCGAAGAGAGAGAGCGCTGCGCCAGCTGAGCGGGCGGAGCAGAGAGAGAGCG 1422  
 QY 1443 GGAGCGCCAGAGAGAGAGCGCAAGCGCAGCTGAGCTGAGAGAGAGAGAGAGCA 1502  
 Db 1423 TGAAGCGCAGAGAGAGAGAGCGCAAGAGAGCACTGAGACTGAGAGAGAGAGCA 1482  
 QY 1503 GCGGAGAGTGGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1562  
 Db 1483 GCGGAGAGTGGAG 1542  
 QY 1563 GCGCAAG 1622  
 Db 1543 TGAAG 1602  
 QY 1623 ACTCCTGAATCAG 1682  
 Db 1603 ACTACTAATCAAG 1562  
 QY 1683 GACTCTGAGAGTGGAG 1742  
 Db 1663 GACTTGGAGATTTGAGATTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1722  
 QY 1743 TCAGAGATTCAGAGTGTGAGCTGCGCAACCCAGAGAGAGAGAGAGAGAGAGAG 1802  
 Db 1723 TCAAGATATCAGATGTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1782  
 QY 1803 TAGAGAGTGAAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1862  
 Db 1783 TAGAGAGTGAAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1842  
 QY 1863 AATGCTTGAAGAGTATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1922  
 Db 1843 AATGCTTGAAGAGTATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1902  
 QY 1923 GCGAGAGAGTGTGAG 1982  
 Db 1903 GCGAGAGAGTGTGAG 1962  
 QY 1983 GCTGCGCGGAG 2042  
 Db 1963 ACTAGCTTGGAG 2022  
 QY 2043 GCTGCGAGAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2102  
 Db 2023 ACTAGAGAGATTTGAG 2082  
 QY 2103 ACAGCACTTCCAG 2151  
 Db 2083 GCAACCACTCCAG 2131

RESULT 14  
 AA234574  
 ID AA234574 standard; cDNA: 3231 BP.  
 AC AA234574;  
 AC  
 AC  
 DT 01-FEB-2000 (first entry)  
 XX  
 DE Human SH3D1A cDNA clone 9.  
 XX  
 KW SH3D1A gene; human; Down's syndrome; leukemia; cancer;  
 KW megalocytic abnormality; myeloproliferative disorder;  
 KW platelet disorder; neural disorder; thrombocytopenia;  
 KW haematopoietic disorder; cognitive dysfunction; microcephaly;  
 KW lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy;  
 KW ss.  
 OS Homo sapiens.  
 XX  
 XX  
 XX  
 FT Key Location/Qualifiers  
 CDS 2..1927

FT /\*tag= a  
 XX MO9953062-A2.  
 XX 21-OCT-1999.  
 XX  
 XX 16-APR-1999; 99MO-US08371.  
 XX 16-APR-1998; 98US-0082007.  
 XX  
 XX (CEDA-) CEDARS SINAI HEALTH SYSTEM.  
 XX korenberg JR, Chen X;  
 XX MPI; 1999-633829/54.  
 XX P-PDB; AAY32158.  
 XX  
 XX Nucleic acid from the human SH3D1A gene and its products, useful for  
 PT the diagnosis and treatment of myeloproliferative disorders and  
 PT leukaemia  
 XX  
 PS Claim 2; Fig 14; 99p; English.

CC This is the nucleotide sequence of a non-full-length cDNA (clone  
 CC 9) corresponding to a novel human SH3 gene, termed the SH3D1A  
 CC gene, that contributes to the development of platelets and the  
 CC pathogenesis of leukaemias, both in general and in particular those  
 CC involving the megakaryocytic lineage. The SH3D1A gene maps to the  
 CC small candidate region for low platelets on chromosome 21.  
 CC Sequencing of 5 different sizes of cDNA clone from foetal brain  
 CC (see A4234570-74) suggests that at least 3 isoforms exist. The  
 CC invention provides methods for the diagnosis and treatment of  
 CC megakaryocytic abnormality, myeloproliferative disorder, platelet  
 CC disorder, acute leukaemia, neural disorders, thrombocytopenia,  
 CC platelet disorder on chromosome 21, low platelets in deletion for  
 CC 21, association of gains in chromosome 21 with leukaemias, neural  
 CC abnormalities, dysfunctions and disorders including brain  
 CC malformations and corresponding cognitive dysfunctions,  
 CC microcephaly, lissencephaly, and colpocephaly. Methods are also  
 CC provided for: suppressing cells unable to regulate themselves;  
 CC screening for a somatic alteration in the SH3D1A gene; monitoring  
 CC the progress and adequacy of a treatment; monitoring tumor risk  
 CC progress or megakaryocytic abnormality, myeloproliferative disorder,  
 CC haemopoietic disorder, platelet disorder or leukaemia; and  
 CC treatment of a subject (including a prenatal subject) having  
 CC platelet disorder, myeloproliferative disorder,  
 CC platelet disorder, leukaemia or neural disorder using a  
 CC nucleic acid that expresses SH3D1A or its antisense nucleic acid.

CC Sequence 3231 BP; 1004 A; 721 C; 712 G; 794 T; 0 other;

Query Match 27.4%; Score 1395.4; DB 20; Length 3231;  
 Best Local Similarity 70.7%; Fred. No. 0;  
 Matches 2359; Conservative 0; Mismatches 506; Indels 470; Gaps 19;

QY 1764 GGCACCCCAAGGAGCAAGATTGAGACCAAGCAAGTCTAGAGAGTAAAGATTCTGA 1823  
 Db 1 GACCAACCAAGGAGCAAGATTGAGACCAAGCAAGTCTAGAGAGTAAAGATTCTGA 60  
 QY 1824 AATCACCCTTACAGCAGCAGAGTTCAGCAATCTCAGCAATCTTGAAGATTCTTC 1883  
 Db 61 AATCACCCTTACAGCAGCAGAGTTCAGCAATCTCAGCAATCTTGAAGATTCTTC 120  
 QY 1884 AGAGAAACAGATCTCGTACCGCTTAAACCAAGTCACACAGCAAGCTTGCATAGAG 1943  
 Db 121 AGAGAAACAGATCTCGTACCGCTTAAACCAAGTCACACAGCAAGCTTGCATAGAG 180  
 QY 1944 CTCGCTTCTTACCTCAAAAAGAGCTTGAAGCAAAAGAGCTGACCGGAGCAGAGCTCCG 2003  
 Db 181 TTCACTTGTAACTTAAAGAGCTTGAAGCAAAAGAGCTGACCGGAGCAGAGCTCCG 240  
 QY 2004 GGAGCAGCTGAGCAGAGTGAAGAGAGCAGCTCAAGCTGAGAGATTATGTTT 2063

Db 241 AGACCACTGATGATGAG 300  
 QY 2064 CAACACAGCTGAAG 2123  
 Db 301 CAATATATAGCTGAAG 360  
 QY 2124 GTCCCTGAGAGCAGCGGAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2183  
 Db 361 GTCCATGAGAGCTGAAG 420  
 QY 2184 GAAGCAAAAG 2243  
 Db 421 AAACAAAG 480  
 QY 2244 TGTGAGCAGAGAG--GAGCAGCAGCGCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2300  
 Db 481 TGTGAGCAG 540  
 QY 2301 GAGGAG 2360  
 Db 541 AAGGAG 600  
 QY 2361 CAAGCAG 2420  
 Db 601 CAAGCTGAG 660  
 QY 2421 CTGCTTACCAAG 2480  
 Db 661 CTGCTTACCAAG 720  
 QY 2481 ATATTACGAG 2540  
 Db 721 GATTATCCGAG 780  
 QY 2541 AGAGATATAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2600  
 Db 781 AGAGATATAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840  
 QY 2601 GCTGAG 2660  
 Db 841 ATTTAAAG 900  
 QY 2661 GATTCCCACTCCAG 2720  
 Db 901 GATTCCCACTCCAG 960  
 QY 2721 TCTGCTGAG 2780  
 Db 961 CTGCTGAG 1020  
 QY 2781 CAAGTGGGAG 2840  
 Db 1021 TAAGTGGGAG 1080  
 QY 2841 CAAGTGGGAG 2900  
 Db 1081 TAAGTGGGAG 1140  
 QY 2901 GAG 2960  
 Db 1141 GAG 1200  
 QY 2961 CCAGGAG 3020  
 Db 1201 CCAGGAG 1260  
 QY 3021 AGACAACTTAATTTTAAAG 3080  
 Db 1261 AGACAACTTAATTTTAAAG 1320  
 QY 3081 GTGAGGTTGAG 3140  
 Db 1321 GTGAGGTTGAG 1380



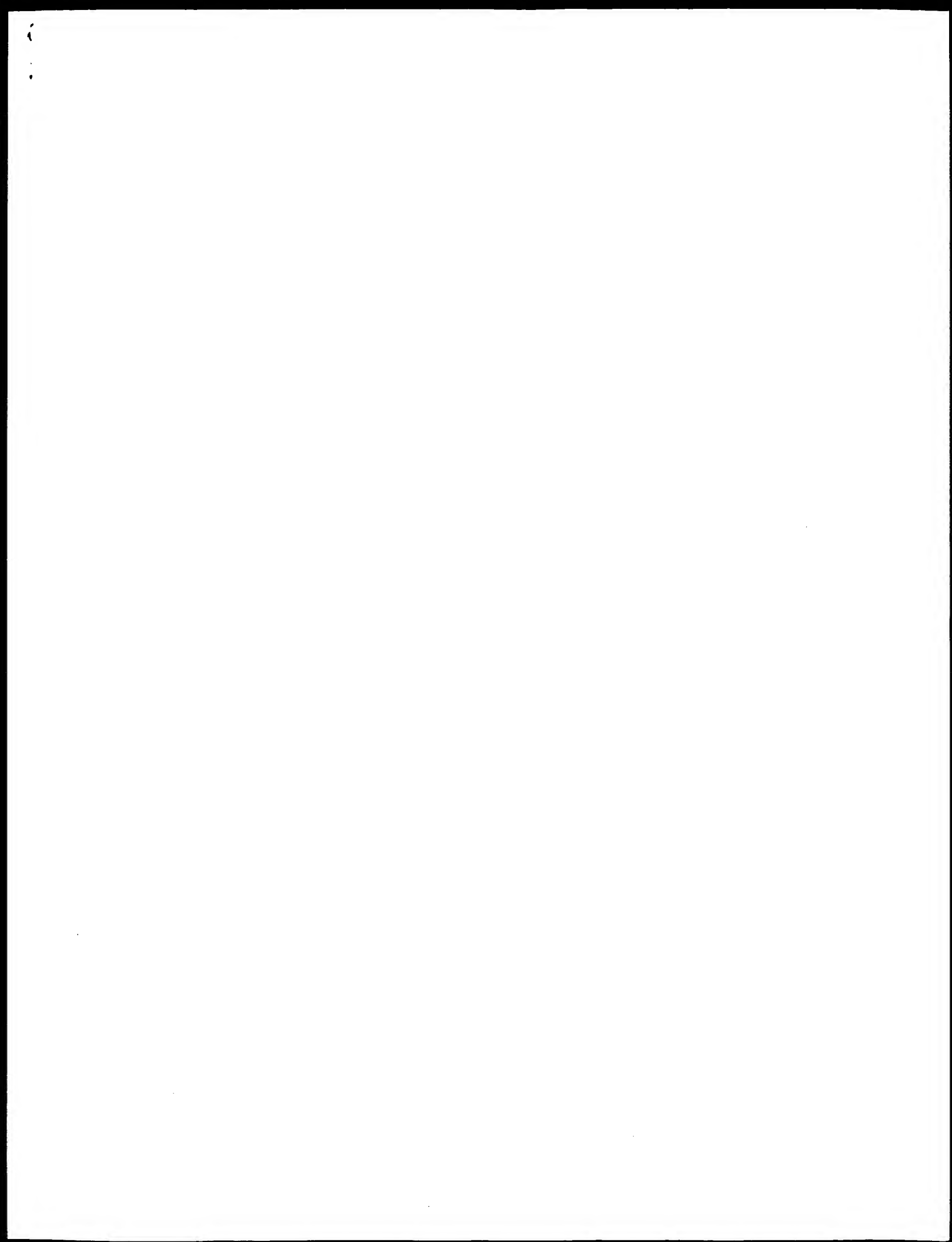




XX Homo sapiens.  
OS  
XX MO200175067-A2.  
PN  
XX 11-OCT-2001.  
PD  
XX 30-MAR-2001; 2001WO-US08631.  
PF  
XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
PI  
XX WPI: 2001-639362/73.  
DR P-PSDB; ABG20575.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
XX Claim 1; SEQ ID No 20566; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations in  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIFO  
CC at ftp.wipo.int/pub/published\_pat\_sequences.  
XX  
XX Sequence 2874 BP; 940 A; 679 C; 703 G; 552 T; 0 other;  
SQ  
Query Match 25.4%; Score 1291.4; DB 23; Length 2874;  
Best Local Similarity 76.0%; Pred. No. 1e-303;  
Matches 1778; Conservative 0; Mismatches 266; Indels 297; Gaps 3;  
1561 GCCGCAAAACGGGAATCTGAAAGCAGCACTTGAATGGAACGGAACCGAGACAG 1620  
Db 1 GCTGCAAAACGGGAATCTGAAAGCAGCACTTGAATGGAACGGAATCGAAGCA 60  
QY 1621 GAACTCTGTAATCAGAGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680  
Db 61 GAACTCTGTAATCAGAGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
QY 1681 AAGACTCTGAGTTGAGTGAAGAGCTCTGAATGACAAAAGATCAGTACAGAG 1740  
Db 121 AAGACTCTGAGTTGAGTGAAGAGCTCTGAATGATTAAGATCACTAGAGAGAA 180  
QY 1741 CTTCAGAGATTCAGGTGTCAGTGGCAACCGAGAGAGAGAGAGAGAGAGAG 1800  
Db 181 CTTCAGAGATTCAGGTGTCAGTGGCAACCGAGAGAGAGAGAGAGAGAGAG 240  
QY 1801 TCTAGAGAGCTAAGATTTGCTGAATCAGCACTTACAGAGAGAGAGAGAGAG 1860  
Db 241 TCTAGAGAGCTAAGATTTGCTGAATCAGCACTTACAGAGAGAGAGAGAGAG 300

QY 1861 CAAATGCTTGGAGACTTATTCAGAGAGAAACAGATCTCAGTACAGCAGTTAAAAAAGTC 1920  
Db 301 CAAATGCTTGGAGACTTATTCAGAGAGAAACAGATCTCAGTACAGCAGTTAAAAAAGTC 360  
QY 1921 CAGCAGAAACAGTTTGCATAGAGACTGCTTCTTAACCTCAAAAAGAGCTTTGAGAGCAAG 1980  
Db 361 CAGCAGAAACAGTTTGCATAGAGACTGCTTCTTAACCTCAAAAAGAGCTTTGAGAGCAAG 420  
QY 1981 GAGCTGCGCCCGCAGAGAGCTCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040  
Db 421 GAGCTGCGCCCGCAGAGAGCTCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
QY 2041 AAGCTGAGAGATTTGATTTGTTTCAACACAGCTGAGAGAGAGAGAGAGATACATAGC 2100  
Db 481 AAGCTGAGAGATTTGATTTGTTTCAACATATCAGCTGAGAGAGAGAGAGATACATAGC 540  
QY 2101 AAACAGCACTCCAG 2160  
Db 541 AAACAGCACTCCAG 600  
QY 2161 GAG 2220  
Db 601 GAG 660  
QY 2221 AGGAGCAAGCAATGCTGAGAGATGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2277  
Db 661 AGGAGCAAGCAATGCTGAGAGATGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720  
QY 2278 CCCACAG 2337  
Db 721 CTCACAG 780  
QY 2338 AGAGCAAG 2397  
Db 781 AGAGCAAG 840  
QY 2398 GCTAAGCTGAG 2457  
Db 841 GCTAAGCTGAG 900  
QY 2458 GCACAG 2517  
Db 901 GCACAG 960  
QY 2518 CACGATGATCAGATCAGATCAGATCAGATCAGATCAGATCAGATCAGATCAGAT 2577  
Db 961 CACGATGATCAGATCAGATCAGATCAGATCAGATCAGATCAGATCAGATCAGAT 1020  
QY 2578 GAGCAG 2637  
Db 1021 GAGCAG 1080  
QY 2638 GCAG 2697  
Db 1081 GCAG 1140  
QY 2698 TCTGCGCCCTGCGCCCAAACTGAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2757  
Db 1141 TCTGCGCCCTGCGCCCAAACTGAGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
QY 2758 TCTGAGCCCTCAG 2817  
Db 1201 TCTGAGCCCTCAG 1260  
QY 2818 TCAACAG 2877  
Db 1261 TCAACAG 1320  
QY 2878 GTACATAGTGTGCGCAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2937  
Db 1321 GTACATAGTGTGCGCAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380





Qy	286	GGTAGGCTGGAGTCTCTGGGCGCATATACGTGTGAGAGAAAGGGCCAGATGATCACGAGATTC	34
Db	69	GGAGGGCCAAATATTTGGGCTATTATACATCTGAAAGAACGTATGAGATGATTAACAGTTT	122
Qy	346	CTTAGGCTGGAAGCCGATAGCGGAAATTTATTACTGTGATCAAGCAGGAACCTTTTCTTC	405
Db	129	GATTAACCTCAAACTTTCAGAGAGTTTACATAAACGGTGATCAAGCCCGTACTTTTCTTA	185
Qy	406	CAATCGGGTTAACCTCAGCCGTGCTTAGCACAATATGAGGCGCTAGCGGACACTGAATTAC	465
Db	189	CAGTCAAGGCTTCGGGCGCCGGGTTTTAGCTGAAATATGGGCCCTTATCAGATCTGAACAG	245
Qy	466	GATGAAAGATGATCAAGTGGAATTTTCCATATGAGCCATGAAGCTTATCAACTGAAGCTA	525



APPLICATION NUMBER: US/09/879,957  
 FILING DATE: 13-Jun-2001  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/630,915  
 FILING DATE: 03-APR-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Mistrock, S. Leslie  
 REGISTRATION NUMBER: 18,872  
 REFERENCE/DOCKET NUMBER: 1101-174  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 790-9090  
 TELEFAX: (212) 869-8864/9741  
 TELEX: 66141 PENNIE  
 INFORMATION FOR SEQ ID NO: 193:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2873 bases  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: unknown  
 MOLECULE TYPE: DNA  
 SEQUENCE DESCRIPTION: SEQ ID NO: 193:  
 US-09-879-957-193

Query Match 6.6%; Score 333.2; DB 10; Length 2873;

Best Local Similarity 61.0%; Pred. No. 3.3e-76;  
 Matches 633; Conservative 0; Mismatches 363; Indels 42; Gaps 4;

2867 CTTCTGTACGCGTACCTGAGTGTGGCCAGTTAGGGCAGATCAGCCTTACCCCGACCA 2926  
 541 CTTTTCATAACCTTACTCTTAATATCATGCGAAGAAAAATCAGCTTCATCTGAACTG 600  
 2927 CAGCCACTGCTCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 2986  
 601 TGCCCTGG---ATCGTATCACTTATTCATGACAGGAGCAAGGTAGAAACTTAA 657  
 2987 AAGCGCAGCCCTGTATCTCTGGAGCCCAAAAAGCAACCACTTAATTTTAAACAAA 3046  
 658 AACGACAGGCTTTGTTCTCGACCTGCAAGAAAGATTAACCACTTGAACTTCTCAAA 717  
 3047 GTGACGTCATCAGCGTCTGGAAGAGCAAGCATGTGGTGTGGAGAACTCAAGGTC 3106  
 718 ATGACATTATTTACTGTTTGGAGCAGCAAGAAATTTGGTTGGAGGTGCAAGAG 777  
 3107 AGAAGGTTGTTCCCAAGTCTTACGTGAAGCTCATTTTCAGGGCCCTTAAGGAAATCCA 3166  
 778 GAAGAGATGTTTCCCAATCTTATGTCAAGATCATTTCTGGAGTGAAGTAAACGGG 837  
 3167 CAAGCATGATCTGCGCCCTACTGAAGCTCTGCTAGTCTTAAAGAGTGGCTTCCCGG 3226  
 838 AA-----GAACGAGAAAGCTTTGTATGACGCTGAATAAAGAACTTACCTCGG 885  
 3227 CCGCAAGCCAGCCATTCGCGGAGAAAGTTTATTTGCAATGATACATACGAAGTTCTG 3286  
 886 CAGCTTATTAAGT-----TGGAGAAAGATATATGACATTTTCAATTTCAAGTGTG 939  
 3287 AGCAAGAGATTTTAACTTTCAGCAAGGGATGATTTGGTTTACCAAGAAAGATGATG 3346  
 940 AACCTGAGATTTTGACCTTTCAGCAAGAGTGAAGAAATTTGGTGAACCCAGAAAGATG 999  
 3347 ACTGTGAGCAGGAACGCTGGGCGACAAAGTCCGAGTCTTCCCTTCTAATATGTAGAG 3406  
 1000 AGTGTGAGCAGGAAGATTTGGAGATAGAGTGAAGTGAATTTTTCATCAAACTATGTCAA 1059  
 3407 TTAAAGATTCAGAGGGCTCTGGAACCTCTGGAAACAGGGAAGTTTAAAGAAAACTTG 3466  
 1060 CAAAGATCAAGAGAGTTTGGAGGTGTCAGCAAGTCTGAGCATCAAAATTAATAAACTG 1119  
 3467 AAATTTGCCAGGTATTTGTTCTTCAAGCTCTACTGTTCCGAAACCACTGCTGCTC 3526  
 1120 AGATTTGCTGAGTACTTACAGCATATGTGCTTGTGTTTCTGAACCAACTTACCTTG 1179  
 3527 CTGGGAGCTGATTTGTATCCGGAAGAAACCAAGGTGATGTTGGAGAGGAACTG 3586

1180 CAGGCACTTATATTAATTTTAAAGAAAAATACAGGGGTGGTGGCAAGAGATTAC 1239  
 3587 AAGCTGAGGAAAAAAGCCAGATAGGTGTTTCCAGCAATATATGCTAACTTTAA 3646  
 1240 AGCCAGAGGAAAAAAGGACAGAAAGATGTTTCTCCGACATGTTAACTTTGG 1299  
 3647 GCCCGGAACAGCAAAATCAACCCCACTGAGCTAACCAAGCCGAGTGCAGCCAG 3706  
 1300 GTCCAAAGTAGMAAGAGCCACACTGCTTTTCATC-----CTG 1338  
 3707 TGTGCAAGTATGAGTGGATGTAAGATTAACCCGCCAAGCATGAGAACTACCTTCA 3766  
 1339 TATGTCAAGTATGCTATGATGATGATGATGATGATGATGATGATGATGATGATG 1398  
 3767 GCAAGGCCAGATATCAACCTCTCAACAGAGAGACCCGAGTGTGGAAGAGAAAG 3826  
 1399 CCAAGGACCACTATTAATTTATGAAACAAAGATGATCTGATTTGGTGCAGAGAGA 1458  
 3827 TCAGTGGCAAGTGGGCTCTTCCCATCAATTTATGTAAGCTGACACAGATGAC 3886  
 1459 TCAAGGGGTGACTGCTCTTCTTCAACTAGCTTAAGATGACAGACATCAGATC 1518  
 3887 CCAGCCAGCAATGAATCA 3904  
 1519 CAAGTCAACAGTGACCA 1536

# RESULT 3

US-09-764-868-125  
 ; Sequence 125; Application US/09/64868  
 ; Patent No. US2002016871A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: PT232  
 ; CURRENT APPLICATION NUMBER: US/09/764,868  
 ; CURRENT FILING DATE: 2001-01-17  
 ; Prior application data removed - refer to PALM or file wrapper  
 ; NUMBER OF SEQ ID NOS: 1510  
 ; SOFTWARE: Patentlin Ver. 2.0  
 ; SEQ ID NO 125  
 ; LENGTH: 4210  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 US-09-764-868-125

Query Match 6.5%; Score 331.4; DB 9; Length 4210;  
 Best Local Similarity 61.0%; Pred. No. 1.2e-75;  
 Matches 630; Conservative 0; Mismatches 361; Indels 42; Gaps 4;

2867 CTTCTGTACGCGTACCTGAGTGTGGCCAGTTAGGGCAGATCAGCCTTACCCCGACCA 2926  
 918 CTTTTCATAACCTTACTCTTAATATCATGCGAAGAAAAATCAGCTTCATCTGAACTG 977  
 2927 CAGCCACTGCTCCTCCCATCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 2986  
 978 TGCCCTGG---ATCGTATCACTTATTCATGACAGGAGCAAGGTAGAAACTTAA 1034  
 2987 AAGCGCAGCCCTGTATCTCTGGAGCCCAAAAAGCAACCACTTAATTTTAAACAAA 3046  
 1035 AAGCAGAGCCCTTTGTTCTCGACCTGCAAGAAAGATTAACCACTTGAACTTCAAAAC 1094  
 3047 GTGACGTCATCAGCGTCTGGAAGAGCAAGCATGTGGTGTGGAGAACTTCAAGGTC 3106  
 1095 ATGACATTATTTACTGTTTGGAGCAGCAAGAAATTTGGTGTGGAGAGTGCATGAG 1154  
 3107 AGAAGGTTGTTTCCCAAGCTTACGTGAAGTCAATTTTCAAGCCCTGTAAGAAATCCA 3166  
 1155 GAAGAGATGTTTCCCAATCTTATGTCAAGATCATTTCTGGAGTGAAGTAAACGGG 1214  
 3167 CAAGCATGATCTGCGCTTCTGTAAGTCTGCTAGTCTTAAAGAGTGGCTTCCCGG 3226



```

Db 1215 AA-----GAAACAGAGCTTTGTATGACGTGTAATTAAGAACTTAACCTCGG 1262
QY 3227 CCGGCAAGCCAGCATTCGCCGAGAGAGTTTATTTGTCATGTAACATACAGAGTTCTG 3286
Db 1263 CAGCTTATTCAGT-----GAGAGAGATATATATGACATTTATCCATATTCAGAGTCTG 1316
QY 3287 AGCAGAGATTTAACCTTTTCAGAGAGGAGTGTATTTGGTTTACAGAGAAAGATGGTG 3346
Db 1317 AACCTGAGATTGACTTTCACAGAGAGTGAAGAAATATTGGTGAGCCCAAGAAAGATGGAG 1376
QY 3347 ACTGTTGAGACGGGAGAGGTGGGCGACAAGTCGGAGTCTTCCTTCACTATGTTAGGC 3406
Db 1377 AGTGTGACAGGAGATTGAGAGATGAAAGTGAATTTTTCATCAACTATGTCAAC 1436
QY 3407 TTAAGATTGAGAGGCTCTGGAACCTGCTGGGAAACAGGAGATTGAGAAAAAACTTG 3466
Db 1437 CAAGAGATCAAGAGATTGTTGGAGTCTGACAAAGTCTGGAGCATCAAAATAAAAACCTG 1496
QY 3467 AAATTGCCAGTTATTGCTTCTACGCTGTACTGTCTCCGAAACAACCTACCTGGCTC 3526
Db 1497 AGATTGCTCAGTTAACTTCAGCATATGTGTCTGTTCTGAAACAATTAGCCTTGGAC 1556
QY 3527 CTGGGAGCTGATTCGTATCCGAGAAAGAAACCCAGGTGATGTTGGAGAGAGAACTGC 3586
Db 1557 CAGAGACATTATATTTATTTTAAAGAAAAATCAAGTGGGTGGTGGCAAGAGAGTTAC 1616
QY 3587 AAGCTCAGAGGAAAAAGCCGAGATAGGGTGTGTTCCAGCAATTTATGCTAACTTTAA 3646
Db 1617 AGGCTCAGAGGAAAAAGCCGAGAGAGATGTTTCTGCGCAGTCAATGTTAACTTTGG 1676
QY 3647 GCGCCGGAACAAGCAAAATATCCCAACTGAGTACCCAGAACCGGAGTGCAGCCAGCAG 3706
Db 1677 GTCCAAAGTAGAGAAAGAGCCACACTCTCTTTTCATC-----CTG 1715
QY 3707 TGTGCCAGGTGATGGGATGTAGCATTAACACCGCCCAAGCATGAGAACTAGCCTTCA 3766
Db 1716 TATGTCAAGTATGTTATGTATGATGACTATGAGCAAAATTAATGAAGATGAGCTCAGTTCT 1775
QY 3767 GCAAGGCGCATATCATCAAGTCTCTCAACAGAGAGACCCGAGTGGTGAAGAGAGAAAG 3826
Db 1776 CCAAGGAGCAACTATTAAATGTATGAACAAAGATGATCTGATGGTGGCAAGAGAGA 1835
QY 3827 TCATGGGCGAGTTGGGCTCTCCCATTCATTCATTAATGTAAAGCTGACCAAGATGGACC 3886
Db 1836 TCAAGGGGTGACGTGCTCTCTTCTTCAAACTAGCTTAAGATGACGACGATCAGATC 1895
QY 3887 CCAGCCAGCAATG 3899
Db 1896 CAAGTCAACAGTG 1908

```

## RESULT 4

US-09-879-957-39  
 Sequence 39, Application US/09879957  
 Patent No. US20020034755A1  
 GENERAL INFORMATION:

APPLICANT: SPARKS, Andrew B.  
 HOFFMAN, No. US20020034755A1h  
 KAY, Brian K.  
 FOWLES, Dana M.  
 McCONNELL, Stephen J.

TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL  
 DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
 USING SAME

NUMBER OF SEQUENCES: 227  
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA

ZIP: 10036-2711  
 COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/879,957
FILING DATE: 13-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/630,915
FILING DATE: 03-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Mistock, Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 747 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 39:
US-09-879-957-39
Query Match 5.1%; Score 258; DB 10; Length 747;
Best Local Similarity 63.9%; Pred. No. 5,4e-57;
Matches 418; Conservative 0; Mismatches 215; Indels 21; Gaps 1;
QY 3247 GGAGAGAGTTTATTTGTCATGTAACATACAGAGTTCTGAGCAAGAGATTAACTTT 3306
Db 115 GGAGAGAGATATTTGTCATTTATTCATATTTCAAGTGTGAACCTGAGATTACTTTG 174
QY 3307 CAGCAAGGAGATGATTTGTTGTTACCAAGAAAGTGTGACTGTGACCGGAAACGCTG 3366
Db 175 ACAGAAAGTGAAGAAATTTGTGTGACCCAGAAAGATGAGAGTGTGACAGAAAGATT 234
QY 3367 GCGGCAAGTCGCGAGTCTCTCTTCTTACTATGTAGAGCTTAAAGATTGAGAGGCTCT 3426
Db 235 GGAGATGAAGAGCAATTTTCCATCAAACTATGTCAAAACCAAGAGATCAAGAGATTT 294
QY 3427 GGAAGTCTGGGAAACAGGAGATTGAGAAAAAACCCTGAATTTGCCAGGTTATTGCT 3486
Db 295 GGAAGTCTGAGCAAGTCTGAGACATCAATTAATAAAACCTGAGATTGCTCAGGTAATT 354
QY 3487 TCCTAGCTGTACTGTGTCGCGAAACACTCACTGCTCTCTGGGAGCTGATTCTGATC 3546
Db 355 GCATATGTTGCTTCTGTTCTGAACTTATGCTTGGACAGAGATTAATTTATT 414
QY 3547 CGGAAAAAGAACCCAGGTGATGTTGAGAGAGAGATGCAAGCTCGAGGAAAAAGCC 3606
Db 415 CTAAAGAAAAATACCAAGTGGGTGGGCAAGAGATGACAGGCAAGAAAAACGA 474
QY 3607 CAGATAGGCTGTTTCCAGCAATATATGTCAAACTTCAAGCTCCCGGAACAGCAAAATC 3666
Db 475 CAGAAAGAGTGTCTTCTGCGCAGTCATTTAACTTTGGGTCCAGAGCTGAAAGAGCC 534
QY 3667 ACCCAACTGAGCTAACCCAGAGACCGCAGTGCAGCCAGAGAGTGCAGAGTATCGGAGTG 3726
Db 535 ACACCTGCTTTCATCT-----GATGTCAAGTATGCTATG 573
QY 3727 TACGATTACAGCCGCAAGAGATGAGCAACTAGCTTACAGAAAGCCAGATCATCAAC 3786
Db 574 TATGACTATGACGAAATAAAGATGAGTCAAGTTTCTCAAGGAGCAACTATTAAT 633
QY 3787 GTCTCAACAGAGAGACCCGAGTGTGAGAAAGAGAGTCACTGGGCAAGTTGGGCTC 3846
Db 634 GTTATGAACAAAGATCTCTGATTGGTGGCAGAGAGAGATCAACGGGGTGAAGTCTGCTC 693

```



```

PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO: 26948
LENGTH: 286
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000311.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.87
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 32
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.66
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.69
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.92
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.98
OTHER INFORMATION: NT HIT: AF114488.1, EVALUE 0.00e+00
OTHER INFORMATION: SWISSPROT HIT: Q15811, EVALUE 2.00e-45
OTHER INFORMATION: EST_HUMAN HIT: AA773263.1, EVALUE 1.00e-112
US-09-864-761-26948

```

```

Query Match 3.9%; Score 198.8; DB 10; Length 286;
Best Local Similarity 87.2%; Pred. No. 7.9e-42;
Matches 218; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```

```

QY 2557 GTGGATGAAAGCCAGACTGAGAGCCAGATGCTTGGAGAGAGCTGAAAGGAAAGAGC 2616
DB 1 GTGGATGAAAGCCAGACTGAGAGCCAGATGCTTGGAGAGAGCTGAAAGGAAAGAGCA 60
QY 2617 GGATGCTTCCCTGCAAACTATGAGAAAAGATCCAGAAAATGAGGTTCCCACTCCAGCC 2676
DB 61 GGGTGGTTCCTGCAAACTATGAGAAAATCCAGAAAATGAGGTTCCCGCTCCAGTG 120
QY 2677 AAACGATGACCGATGACATCTGCTGCCCCCTGCCCCCAAACTGCTGCTGAGACCCCT 2736
DB 121 AAACGATGACCGATGACATCTGCTGCCCCCTGCCCCCAAACTGCTGCTGAGACCCCT 180
QY 2737 GCTCCTTTGGCAGTACCTCTTGTAGCCCTCCAGAACCCCAACGACTGGGCGAGACTTC 2796
DB 181 GCCCTTTGGCAGTACCTCTTGTAGCCCTCCAGAACCCCAACGACTGGGCGAGACTTC 240
QY 2797 AGTTCCACGT 2806
DB 241 AGTTCCACGT 250

```

RESULT 7

```

US-09-864-761-30453
Sequence 30453, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmics-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366

```

```

PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 30453
LENGTH: 297
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000117.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.7
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95
OTHER INFORMATION: NT HIT: AF114487.1, EVALUE 0.00e+00
OTHER INFORMATION: EST_HUMAN HIT: AA773263.1, EVALUE 1.00e-112
OTHER INFORMATION: SWISSPROT HIT: Q15811, EVALUE 2.00e-45
US-09-864-761-30453

```

Query Match 3.9%; Score 198.8; DB 10; Length 297;

Best Local Similarity 87.2%; Pred. No. 8.1e-42;

Matches 218; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```

QY 2557 GTGGATGAAAGCCAGACTGAGAGCCAGATGCTTGGAGAGAGCTGAAAGGAAAGAGC 2616
DB 1 GTGGATGAAAGCCAGACTGAGAGCCAGATGCTTGGAGAGAGCTGAAAGGAAAGAGCA 60
QY 2617 GGATGCTTCCCTGCAAACTATGAGAAAAGATCCAGAAAATGAGGTTCCCACTCCAGCC 2676
DB 61 GGGTGGTTCCTGCAAACTATGAGAAAATCCAGAAAATGAGGTTCCCGCTCCAGTG 120
QY 2677 AAACGATGACCGATGACATCTGCTGCCCCCTGCCCCCAAACTGCTGCTGAGACCCCT 2736
DB 121 AAACGATGACCGATGACATCTGCTGCCCCCTGCCCCCAAACTGCTGCTGAGACCCCT 180
QY 2737 GCTCCTTTGGCAGTACCTCTTGTAGCCCTCCAGAACCCCAACGACTGGGCGAGACTTC 2796
DB 181 GCCCTTTGGCAGTACCTCTTGTAGCCCTCCAGAACCCCAACGACTGGGCGAGACTTC 240
QY 2797 AGTTCCACGT 2806

```

DB 241 AGCTCCACCT 250

```

RESULT 8
US-09-864-761-17146
; Sequence 17146, Application US/09864761
; Patent No. US20020048763A1
GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 17146
LENGTH: 301
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000311.1
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.3
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.7
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2
OTHER INFORMATION: SWISSPROT HIT: O35601, EVALUATE 3.50e-01

```

```

; OTHER INFORMATION: NT HIT: AF114487.1, EVALUATE 1.00e-122
; OTHER INFORMATION: EST_HUMAN HIT: BE542917.1, EVALUATE 4.00e-62
US-09-864-761-17146
Query Match 3.9%; Score 198.6; DB 10; Length 301;
Best Local Similarity 82.3%; Pred. No. 9.2e-42;
Matches 228; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
1298 TGACATTTGAAGATAGAGCGGAGAACTTCGAGCGAGCGAGCTGTGAGCTGAGAGC 1357
1 TAACGTTTGAAGATAGAGCGGAGAACTTTGAACTGCGCAACTGGAAGTGAAGAAC 60
1358 GCCGCCAAGCCCTCTTGGAGCAGCAGCGCAAGCAGCAGCGCTGTGAGCTGAGC 1417
61 GAAGCAGAGCTCTCTGGAACAGCAGCGCAAGCAGCAGCGCTGTGAGCTGAGC 120
1418 GCGCGAGCAGCAGAGAGGAAAGCGGAGCGCCAGCAGCAGAGCGCAAGCGGAGCTGG 1477
121 GGGCGAGCAGAGAGAGAGAGCGGTAGCGCCAGGAGCAAGCGCAAGACAACTGG 180
1478 AGCTGAGAGACAGCTGAGAGAGCGGAGCGGAGCTGTGAGCGCAGCAGAGAGAGAGA 1537
181 AACTGAGAGACAGCTGAGAGAGCGGAGCGGAGCTGTGAGCGCAGAGAGAGAGAGAGA 240
1538 GAAGAGAGATGAGAGCGCGCAGCGCCGCAAAACGGCA 1574
241 GAAAGAAATTGAGAGCGCGAGGCTAAGCAGCGCGAGA 277
RESULT 9
US-09-864-761-17644
; Sequence 17644, Application US/09864761
; Patent No. US20020048763A1
GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecmca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30

```

PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 17644  
LENGTH: 263  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AP000049.1  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.3  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.3  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3  
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.9  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.5  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9  
OTHER INFORMATION: EST\_HUMAN HIT: BE542917.1, EVALUE 3.00e-62  
OTHER INFORMATION: SWISSPROT HIT: Q15601, EVALUE 2.80e-01  
OTHER INFORMATION: NT HIT: AF114487.1, EVALUE 1.00e-122  
US-09-674-761-17644

Query Match 3.9%; Score 197.4; DB 10; Length 263;  
Best Local Similarity 84.4%; Pred No. 1.7e-41;  
Matches 222; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 1298 TGACATTTGAAGATAAGAGCGGAGAACTTCGAGCGAGCACTGTGAGCTTGAGAGAC 1357  
DB 1 TAACGTTTGAAGATAAGAGCGGAGAACTTCGAGCGAGCACTGTGAGAGAAAC 60  
QY 1358 GCGCGCAAGCGCTTGGAGCAGCAGCAAGAGAGAGAGCGGCTGCTCAGCTGAGGC 1417  
DB 61 GAAGGCAAGCTTCTCTGGAACAGAGGAGAGAGAGAGAGCGCTGCGCCAGCTGAGGC 120  
QY 1418 GCGCGCAGCAGAGAGAGAGAGAGCGGAGCGCAGAGAGAGAGCGCGAGCTTG 1477  
DB 121 GCGCGAGCAGAGAGAGAGAGAGCGGAGCGCAGAGAGAGAGCGCGAGAGAGAGAG 180  
QY 1478 AGCTGAG 1537  
DB 181 AACTGAG 240  
QY 1538 GGAAGAGATCGAGAGCGCGAG 1560  
DB 241 GGAAGAGATTCGAGAGCGCGAG 263

RESULT 10  
US-09-674-761-10314  
Sequence 10314, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aemica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 10314  
LENGTH: 480  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AP000311.1  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.87  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.66  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.69  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 0.92  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.98  
US-09-674-761-10314

Query Match 3.8%; Score 192; DB 10; Length 480;  
Best Local Similarity 81.6%; Pred No. 6.5e-40;  
Matches 222; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 2523 TGAGATCAACATCCAGCAGAGATATATGATGATGATGAAAGCCAGCTGAGAGCC 2582  
DB 209 TTAGCTTACTTAAGCTGAGACTTTTCCACAGGTGATGAAAGCCAACTGAGAAAC 268  
QY 2583 AGATGCTTGGAG 2642  
DB 269 CGGCTGCTTGGAG 328  
QY 2643 AAGATTCAGAAATAGGTTCCACTCCAGCCAAACAGTAGACCATCTGAGCATCTGC 2702  
DB 329 GAAATTCAGAAATAGGTTCCCGCTCCAGTAAACAGTAGATTCATCAATCTGC 368  
QY 2703 CCTGCCCCCAAGCTGCTGCTGAGAGCCCTGCTCTTTCCAGTAGACCTTCTTGA 2762  
DB 389 CCGTCCCCCAAGCTGCTGCTGAGAGCCCGCTCTTTGAGAGTAACTCTTCA 448  
QY 2763 GCGCTCCAGCAACCCCAAGCACTGGGAGACT 2794  
DB 449 GCGCTCCAGCAACCCCTAATATGAGGCGGACT 480

## RESULT 11

US-09-864-761-333

Sequence 333, Application US/09864761

Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

FILE REFERENCE: Aecmca-X-1

CURRENT FILING DATE: 2001-05-23

CURRENT FILING DATE: 2001-05-23

PRIOR FILING DATE: US 60/180,312

PRIOR FILING DATE: 2000-02-04

PRIOR FILING DATE: 2000-05-26

PRIOR FILING DATE: 2000-05-26

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263.6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00668

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: US 60/234,687

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: US 09/608,408

PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: US 09/774,203

PRIOR FILING DATE: 2001-01-29

NUMBER OF SEQ ID NOS: 49117

SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

SEQ ID NO 333

LENGTH: 486

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: MAP TO AP000311.1

OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2

OTHER INFORMATION: EXPRESSED IN HELI00, SIGNAL = 3.2

OTHER INFORMATION: EXPRESSED IN PETAL LIVER, SIGNAL = 3

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.3

OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.2

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4

OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.7

OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3

OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2

US-09-864-761-333

Query Match

3.6%; Score 180.8; DB 10; Length 486;

Best Local Similarity 84.6%; Pred. No. 5,4e-37; Matches 203; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 1297 GTGACATTGAAGTAAGAACCGGAGAACTTCGAGCGAGCGAGTGTGAGTGTGAGAG 1356

Db 247 GTAACGTTGAAGTAAGAACCGGAGAACTTCGAGCGAGCGAGTGTGAGTGTGAGAG 306

QY 1357 CGCCGCCAAGCGCTTGTGAGCAGCAGCGCAAGAGCAGCGGCTTGTGAGTGTGAG 1416

Db 307 CGAAGCAAGCTCTCTGGAACAGCAGCGCAAGAGCAGCGGCTTGTGAGTGTGAG 366

QY 1417 CGCCGCCAAGCGCTTGTGAGCAGCAGCGCAAGAGCAGCGGCTTGTGAGTGTGAG 1476

Db 367 CGGCGGAGCAGGAGGAGGAGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGG 426

QY 1477 GAGCTGGAAGCAGCTGTGAGCAGCAGCGGAGCTGTGAGCGGAGCGGAGCGGAGG 1536

Db 427 GAATGGAGAGCAACTGTGAAAGCAAGCGGAGCTGTGAGCGGAGCGGAGCGGAGG 486

## RESULT 12

US-09-864-761-864

Sequence 864, Application US/09864761

Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

FILE REFERENCE: Aecmca-X-1

CURRENT FILING DATE: 2001-05-23

CURRENT FILING DATE: 2001-05-23

PRIOR FILING DATE: US 60/180,312

PRIOR FILING DATE: 2000-02-04

PRIOR FILING DATE: 2000-05-26

PRIOR FILING DATE: 2000-05-26

PRIOR FILING DATE: 2000-08-03

PRIOR APPLICATION NUMBER: GB 24263.6

PRIOR FILING DATE: 2000-10-04

PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00664

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00669

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00665

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00668

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00663

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00662

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00670

PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: US 60/234,687

PRIOR FILING DATE: 2000-09-21

PRIOR APPLICATION NUMBER: US 09/608,408

PRIOR FILING DATE: 2000-06-30

PRIOR APPLICATION NUMBER: US 09/774,203

PRIOR FILING DATE: 2001-01-29

NUMBER OF SEQ ID NOS: 49117

SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

SEQ ID NO 864

Query Match



```

: LENGTH: 487
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AP000049.1
: OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.3
: OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.3
: OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.9
: OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.2
: OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.5
: OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3
: OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3
: OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
US-09-864-761-864

Query Match          3.6%; Score 180.8; DB 10; Length 487;
Best Local Similarity 84.6%; Pred. No. 5.4e-37;
Matches 203; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 1297 GTGACATTTGAAGATAGAAAGCGGAGAACTTGACGAGGAGTGTGAGCTGAGAG 1356
DB 248 GTACGTTGAATAGAAAGCGGAGAACTTTGAACGTGGCACTGGAAGTGGAGAA 307
QY 1357 CGCGCCAAAGCGCTCTTGAGAGCAGCGCAAGAGCAGAGCGGTTGCTCAGCTGAG 1416
DB 308 CGAAGCAAGCTCTCTGGAACAGCAGCGCAAGAGCAGAGCGCTGCGCAGCTGGAG 367
QY 1417 CGGCGGAGAGAGAGAAAGCGGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1476
DB 368 CGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 427
QY 1477 GACCTGAGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1536
DB 428 GAACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 487

RESULT 13
US-09-864-761-311
: Sequence 311, Application US/09864761
: Patent No. US20020048763A1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharon G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME- DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
: FILE REFERENCE: Aeomica-X-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: PRIOR FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263.6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668

```

```

: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
: SEQ ID NO 311
: LENGTH: 475
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AP000193.1
: OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.2
: OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.3
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
: OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2
: OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
: OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2
: OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.6
: OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7
: OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
US-09-864-761-311

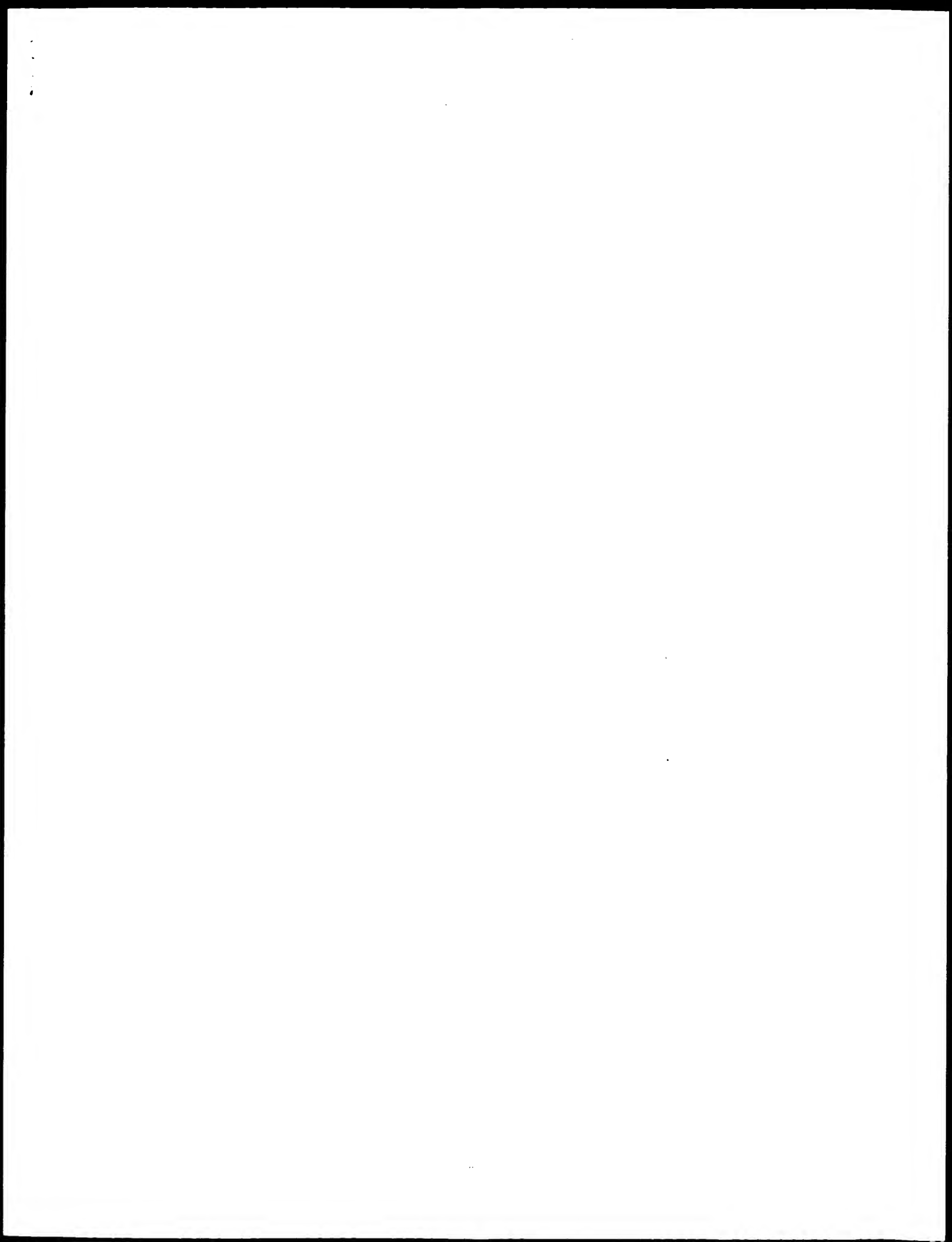
Query Match          3.2%; Score 163; DB 10; Length 475;
Best Local Similarity 80.9%; Pred. No. 2.3e-32;
Matches 190; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 2523 TGACATCACCATCCAGCAGAGATATGTCATGTGTGATGAAAGCCAGCTGAGAGCC 2582
DB 241 TTAGTTCTACTTAAGCTGTGACTTTTCCACAGGTGATGAAAGCCAACTGAGAGACC 300
QY 2583 AGATGCTTGAAGAGAGAGCTGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2642
DB 301 CGGCTGCTTGAAGAGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
QY 2643 AAAGATTCCAGAAATAGAGTTCCACTCCAGCCAAACAGTACCGATCTGACATCTGC 2702
DB 361 GAAATCCAGAAATAGAGTTCCCGCTCCAGTGAACAGTACGATTCATCAATCTGC 420
QY 2703 CCCGCCCCCAAACTGCTGTGCTGAGAGCCCTGCTCTTGCAGTGAACCTCT 2757
DB 421 CCTGCCCCCAAACTGCTGTGCTGAGAGCCCTGCTCTTGCAGTGAACCTCT 475

RESULT 14
US-09-864-761-13884
: Sequence 13884, Application US/09864761
: Patent No. US20020048763A1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharon G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME- DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
: FILE REFERENCE: Aeomica-X-1
: CURRENT APPLICATION NUMBER: US/09/864,761
: PRIOR FILING DATE: 2001-05-23
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456

```





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2003, 01:40:35 ; Search time 618.75 seconds  
(without alignments) 13456.658 Million cell updates/sec

Title: US-09-674-237A-1  
Perfect score: 5084  
Sequence: 1 cggcgcaggagagagagtgag.....gagaatcgatcaagctt 5084

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues  
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
EST :  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hnc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hnc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vit:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1598.2	31.4	2126	11	BC020269 Homo sapi
2	1582	31.1	2079	11	BC013578 Homo sapi
3	795.6	15.6	911	14	BQ942708 AGENCOURT
4	722.8	14.2	732	14	BM950476 UI-M-EHOP
5	721.2	14.2	738	14	BM944544 UI-M-EHOP
6	691.2	13.6	782	13	BG922629 602821583

7	660.2	13.0	896	14	BQ941336	BQ941336 AGENCOURT
8	658.2	12.9	979	14	BQ719508	BQ719508 AGENCOURT
9	656.6	12.9	690	14	BQ179493	BQ179493 UI-M-EHOP
10	633.6	12.5	650	10	BB656585	BB656585 BB656585
11	631.4	12.4	1062	14	BQ898718	BQ898718 AGENCOURT
12	616	12.1	674	9	AV246405	AV246405 AV246405
13	606.8	11.9	610	14	BQ443095	BQ443095 UI-M-EHOP
14	602	11.8	603	9	A1594919	A1594919 m779910.Y
15	593.2	11.7	606	10	AM910367	AM910367 ur80908.Y
16	587	11.5	606	13	BM233613	BM233613 K0339C08-
17	580	11.4	599	13	BM233591	BM233591 K0339A05-
18	580	11.4	599	13	BM247316	BM247316 K0800A10-
19	563.4	11.1	773	14	BM681943	BM681943 UI-E-EOL-
20	558.8	11.0	639	14	BQ201258	BQ201258 UI-R-DQ1-
21	549	10.8	979	14	BQ941411	BQ941411 AGENCOURT
22	541.4	10.6	1089	14	BQ050397	BQ050397 AGENCOURT
23	537.8	10.6	724	10	BE373101	BE373101 601224527
24	533.8	10.5	537	9	AA063751	AA063751 m779910.Y
25	530.4	10.4	966	12	BG297300	BG297300 60239541.8
26	527	10.4	734	9	A1671143	A1671143 w013909.X
27	525.2	10.3	574	13	BM211772	BM211772 C0812B08-
28	522.6	10.3	1064	12	BE888832	BE888832 601513752
29	521.4	10.3	875	12	BF122946	BF122946 601761612
30	517.2	10.2	585	13	BM235301	BM235301 K0416C10-
31	509.2	10.0	576	12	BE949344	BE949344 UI-M-BH3-
32	508.4	10.0	521	9	A1303871	A1303871 ui57d09.Y
33	508.4	10.0	510	12	BE952536	BE952536 UI-M-CEO-
34	501.2	9.9	550	14	BO561212	BO561212 H4068H12-
35	501	9.9	501	9	A1852070	A1852070 UI-M-BHO-
36	498.8	9.8	547	9	A1172344	A1172344 mt04411.X
37	497.8	9.8	597	17	A2422217	A2422217 lW0200L20
38	489.6	9.6	866	10	BE538712	BE538712 601064906
39	484.6	9.5	527	10	BB757493	BB757493 BB757493
40	482	9.5	811	12	BG829540	BG829540 602763842
41	482	9.5	546	12	BF398136	BF398136 UI-R-B52-
42	480.8	9.5	600	9	AL134506	AL134506 DKFZP547M
43	479.8	9.4	483	10	BE200514	BE200514 Ug70a11.Y
44	475.6	9.4	612	12	BG894850	BG894850 355615 MA
45	475	9.3	475	10	BE199657	BE199657 ug70a11.X

## ALIGNMENTS

RESULT 1  
LOCUS BC020269 2126 bp mRNA linear HTC 19-DEC-2001  
DEFINITION Homo sapiens, clone IMAGE:4899011, mRNA.  
ACCESSION BC020269  
VERSION BC020269.1 GI:17939664  
KEYWORDS HTC.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE Strausberg R.  
Direct Submission  
Submitted (19-DEC-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov  
COMMENT Contact: MGC help desk  
Email: cgaabs-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
info@bcgsc.bc.ca  
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,

Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice Meeley, Steven Ness, Pawan Pandoh, Anna-Lisa Parvahu, Parvaneh Saeedi, Jacqueline Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILN at: <http://image.jnl.gov>  
Series: IRAL Plate: 40 Row: n Column: 1  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4504796  
This clone has the following problem: no cloning site / microdeletion.

## FEATURES

Location/Qualifiers  
1. .2126

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:489011"  
/issue\_type="Pancreas, epithelioid carcinoma"  
/clone\_lib="NIH MGC 42"  
/lab\_host="DH10B-R"  
/note="Vector: pOMB7"

BASE COUNT 712 a 462 c 517 g 435 t

## ORIGIN

Query Match 31.4%; Score 1598.2; DB 11; Length 2126;  
Best Local Similarity 85.3%; Pred. No. 0;

Matches 1822; Conservative 0; Mismatches 298; Indels 15; Gaps 3;

```

QY 68 GCAGAGGAGCATCCCGAGCGGCTCCGGAGCGGCCGGAGGCGAGCGGGCGGCG 127
Db 1 GCAGAGGAGCATCCCGAGCGGCTCCGGAGCG-----GACAGAGAGCGCGGCG 49

QY 128 GATGATGATCGCGCGCGCGCATCGGCTTCCTCCGCGCGCGCGCGCGCGCAT 186
Db 50 GATGATGATCGCGCGCGCGCGCATCGGCTTCCTCCGCGCGCGCGCGCGCGCAT 109

QY 187 TTGTGTAGAGGCGCGCGCGCGCGCGCGCGCGCGAGATGAGCGCTGCATGCA 246
Db 110 TTGTGTAGAGGCGCGCGCGCGCGCGCGCGCGCGAGATGAGCGCTGCATGCA 169

QY 247 CGATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 306
Db 170 AGTAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 229

QY 307 ATAACTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 366
Db 230 ATAACTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 289

QY 367 GATTTATTAATGATGATCAAGGAGGAACTTTTTCATCTGGGTTTACCTCA 426
Db 290 GATTTATTAATGATGATCAAGGAGGAACTTTTTCATCTGGGTTTACCTCA 349

QY 427 GTCCTGACCAATATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 486
Db 350 GTCCTGACCAATATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 409

QY 487 GAATTTTCATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 546
Db 410 GAATTTTCATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 469

QY 547 AACTTTCCCCCTGTCTATGAAACAGCAACGATGATTTTTCAGTGCACAGAT 606
Db 470 AACTTTCCCCCTGTCTATGAAACAGCAACGATGATTTTTCAGTGCACAGAT 529

QY 607 ATAGAGGAGGATGCTAGCATGCGACACTCAGAGCTGTGTCTGTGCAATGGAG 666
Db 530 ATAGAGGAGGATGCTAGCATGCGACACTCAGAGCTGTGTCTGTGCAATGGAG 589

QY 667 ATTTCCAGTTGTTGAATGCTCTCAACCTTATGATCTTCTGCTCCAGAGAG 726
Db 590 ATTTCCAGTTGTTGAATGCTCTCAACCTTATGATCTTCTGCTCCAGAGAG 649

```

```

QY 727 CCCCTGGCTAAGGGGCTCTCCCGCATATAGGCTCTGCTGCTTGGCATCTTCA 786
Db 650 CCCCTGGCTAAGGGGCTCTCCCGCATATAGGCTCTGCTGCTTGGCATCTTCA 709

QY 787 GCCATGAGGCAAGAGGATCTTCTTTCAGCAGATCTGTGTCCAGGGTCAATTA 846
Db 710 GCCATGAGGCAAGAGGATCTTCTTTCAGCAGATCTGTGTCCAGGGTCAATTA 769

QY 847 AAGTTACGAAAGGCAATCAATGATGTGCGCAGCGCCCTCCAGCAGAGAAATG 906
Db 770 AAGTTACGAAAGGCAATCAATGATGTGCGCAGCGCTCCAGCAGAGAGTGGCT 829

QY 907 GTCCTGATCATCAAGGCTGAAATATACAGGAGTATTTCAACCCGACGAAATG 966
Db 830 GTCCTGATCATCAAGGCTGAAATATACAGGAGTATTTCAACCCGACGAAATG 889

QY 967 AGTGACACTTAACAGGCTCCCGAGCAAGAACTATTCATGCAATCAAGTTTACC 1026
Db 890 AGTGACACTTAACAGGCTCCCGAGCAAGAACTATTCATGCAATCAAGTTTACC 949

QY 1027 GCTCAGCTGCTTCAATATGGAATCTTTTGCACATTTATCAAGATGGAATCA 1086
Db 950 GCTCAGCTGCTTCAATATGGAATCTTTTGCACATTTATCAAGATGGAATCA 1009

QY 1087 GAAGAAATTTATCTAGCTATGCACTAATGATGTTGCTGCTGCTGCTGCTGCTG 1146
Db 1010 GAAGAAATTTATCTAGCTATGCACTAATGATGTTGCTGCTGCTGCTGCTGCTG 1069

QY 1147 CCGCTGCTGCTTCCAGAAATATATCTCTTCTTCCAGAAAGTTGCTGCGAGTGG 1206
Db 1070 CCGCTGCTGCTTCCAGAAATATATCTCTTCTTCCAGAAAGTTGCTGCGAGTGG 1129

QY 1207 ATGTCCTGATTAAGCTCTTCTGCTGATGATGATGATGATGATGATGATGAT 1266
Db 1130 ATGTCCTGATTAAGCTCTTCTGCTGATGATGATGATGATGATGATGATGAT 1189

QY 1267 GATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1323
Db 1190 GATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1249

QY 1324 AACTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1383
Db 1250 AACTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1309

QY 1384 CGCAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1443
Db 1310 CGCAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1369

QY 1444 GAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1503
Db 1370 GAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1429

QY 1504 CGGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1563
Db 1430 CGGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1489

QY 1564 GCAAAAGGGAAGCTGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1623
Db 1490 GCAAAAGGGAAGCTGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1549

QY 1624 CTCCTGAATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1683
Db 1550 CTCCTGAATGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1609

QY 1684 ACTCTGAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1743
Db 1610 ACTCTGAGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1669

QY 1744 CAGATATACAGGTGTGACTGCGCAACAGAGGCAAGAAATTTAGAGCAAGAA 1803
Db 1670 CAGATATACAGGTGTGACTGCGCAACAGAGGCAAGAAATTTAGAGCAAGAA 1729

```





Qy	1088	AAGAAATTATCCTAGCTATGCAACCTAATTATGTTGCAATGCTGGTACGACCTGCCG	1147
Dp	961	AGGAATTTATCTGGCAATGCACTCTATTGATGAGTAATGTTGGCCAAACCATCTGCCAC	10220
Qy	1148	CCGTCCTCGCTCCAGAAATACATCCCTCTCTTCCAGAAAGATTGCTCCGGCAGTGGGA	1207
Dp	1021	CTGTCCTCGCTCCAGAAATACATCCCTCTTTCAGAAAGATTGCTCCAGTGGGATGA	10868
Qy	1208	TGTCCGTCATTAAGCTCTTCTCTGTGGATCAGAGGCTGCTGAGAGCCGCTGTCTAGAG	1267
Dp	1081	TATCTGTATTAAGCTCAACATCTGTATTCAGAGGCTTACAGAGGGAACCATGTTTAGAG	1140
Qy	1268	ATGAGCAGCAGC -- CAGAGAAAGAACTGCTGTGACATTTGAAAGTAAGACGGGAGA	1324
Dp	1141	ATGAACAACAATATTAGAAAGAAATTCCTGTGAACGTTTGAAAGTTAAGACGGGAGA	1200
Qy	1325	ACTTGACCGAGGCAAGTGTGAGCTGGAGAAAGCCGCCCAAGCGCTTTGGACGACGAC	1384
Dp	1201	ACTTTGAACGTTGGCAACCTTGAACTGGAGAAAGAAAGGCAAGCTCTCTGGAACGACGAC	1260
Qy	1385	GCAAAAGACACAGAGCCGTTGGCTCAGCTGAGCGCCCGACAGAGAGAGAAAGACGGG	1444
Dp	1261	GCAAGAGACACAGAGCCGCTGGCCAGCTGGAGCGGCGGAGCAGAGAGAAAGAGCGTG	1320
Qy	1445	AGCGCAGAGACAGAGGCCCAAGCGGACGTGAGAGCTGGAAGACGCTGGAAAGCAGC	1504
Dp	1321	AGCGCAGAGACCAAGAGCGCAAAAGAACATCTGGAACGTGAGAGAACATCTGGAAGACGACG	1380
Qy	1505	GGGACCTGAGCCGACAGAGAGAGAGAGAGAGAGAGAGAGATCCAGAGGCGCCGAGCGG	1564
Dp	1381	GGGACCTGAGACCGCAGAGAGAGAGAGAGAGAGAGAGAGATTTGAGAGCGCAGAGGCTG	1440
Qy	1565	CAAAAACGGGAACTGAAAAAGCAGCGAACATTTGAGAACGGAACCGGAGACAGAGAC	1624
Dp	1441	CAAAAACGGGAACTTGAAAGGCAACGACAACTTGAGTGGGAAAGGATCGAAGGCAAGAC	1500
Qy	1625	TCTGGAATCAGAGGAACAAGAGAGAGAGAGGACCGTGTCTTGAAGGCACAGAGAAAGA	1684
Dp	1501	TACTAATCAAAAGAAACAAAGAACAGAGGACATGATTTACTTGAAGCAAGAAAGAAAGA	1560
Qy	1685	CTCTGAGTTTGAAGTTGAAAGCTCTGATATACAAAAGCATCAGCTAGAGAGAAACCTTC	1744
Dp	1561	CTTTGGAATTTGAATTTGAAAGCTCTTAATGATTAAGATTAAGATTAAGAGGAAACCTTC	1620
Qy	1745	AGGATATCAGGTGTCCACTGCGCAACCCAGAGCCAGAGAAATTAGAGCACGAACGTCTA	1804
Dp	1621	AAGATATCAGATGTGCGATTACCAACCCAAAGGCAAGAAATTTAGAGCACAAACAACTCTA	1680
Qy	1805	GAGAGCTTAAGATTTGCTGGAATCACCCCATCTAAGCAGAGAGTTGAGAGATCTCAGCAAA	1864
Dp	1681	GAGAGTTGAGAAATTTGCCGAATCATCCATCTACAGCAACAAATTAAGAGAAATCTCAGCAAA	1740
Qy	1865	TGCTTGAAGACCTTATTTCCAGAGAAACAGATACTCAGTACCAAGTTAAACCAAGTCCAGC	1924
Dp	1741	TGCTTGAAGACCTTATTTCCAGAAAAAACAGATCTCAATGACCAATTTAAACCAAGTTGAGC	1800
Qy	1925	AGAACAGTTTGAATAGAGACTGCTGCTTTAACCCTCAAAAGACCTTTGAGACCAAGAGAC	1984
Dp	1801	AGAACAGTTTGAACAGAGATTCACTTTTACATTTAAAGAACCTTTAAGACCAAAAGAAC	1860
Qy	1985	TGGCCCGGACAGCAGCTCCGGAGACAGCTGAGCAGAGGTGAGAGAGAGACAGCTCAAGC	2044
Dp	1861	TAGCTCGGACAGCACTTACAGAGACCAACTGATGTAAATGAGAGAAAGAACTAGATCTAAAC	1920
Qy	2045	TGCAAGAGATTGATGTTTCAACAACCAAGCTGAAGCACTGAGAGATACATAGCAAC	2104
Dp	1921	TACAGGAGATTGATATTTTCAATAATACAGCTGAAGGAACTAAGAGAAATACACAATTAAGC	1980
Qy	2105	AGCAACTCCGAAGCAGAGGCTCCCTGAGAGCAGCGGACTGAAGCAGAGAAACAGCAGAGAGA	2164
Dp	1981	AACCACTCCGAAGCAAAAGTCCATGAGGCTGTAAACGACTGAACACGAAAGAAACAAAGAAC	2040

QY	2165	GGAGAGCCTGGAGTTAGAGAGCAAAAAGGA	2198
Db	2041	GAAGATCATAGATTGAAAAACAAAAAAA	2074
RESULT 3			
LOCUS	BQ942708	911 bp	MRNA linear EST 21-AUG-2002
DEFINITION	AGENCYCOURT 8764676 NIH_MGC_129	Mus musculus	cDNA clone IMAGE:6314690
ACCESSION	BQ942708		
VERSION	BQ942708.1	GI:22358186	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 911)		
TITLE	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a>		
	Tissue Procurement: Susan L. Sullivan, PhD.		
	cDNA Library Preparation: ResGen, Invitrogen Corp		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)		
	DNA Sequencing by: Agencourt Bioscience Corporation		
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: <a href="http://image.lnl.gov">http://image.lnl.gov</a>		
	Plate: LLM13743	row: n	column: 03
	High quality sequence start: 10		
	High quality sequence stop: 564.		
FEATURES	location/Qualifiers		
source	1..911		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/clone="IMAGE:6314690"		
	/clone_id="NIH_MGC_129"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: olfactory epithelium; Vector: pCMW-SORT6.1.cdb; Site 1: ECFV; Site 2: NotI; Cloned directionally. Primer: Oligo dt. Average insert size 2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH-MGC library."		
BASE COUNT	271 a	210 c	280 g 150 t
ORIGIN			
Query Match	15.6%;	Score 795.6;	DB 14; Length 911;
Best Local Similarity	98.0%;	Pred. No. 4.3e-179;	
Matches 890;	Conservative 0;	Mismatches 9;	Indels 9; Gaps 8;
QY	936	GGAGTTATTCAACAGCCACGACAAAACTAGATGAGACACTTAAACAGTCCCGCAGCAAG	995
Db	1	GCAGTTATTCAACAGCCACGACAAAACTATGAGGACACTTAAACAGTCCCGCAGCAAG	60
QY	996	AACATTTCATGCAATGAATTATACCCAGGCTCAGTGGCTTCATATGAAATCTTTC	1055
Db	61	AACATTTCATGCAATGAATTATACCCAGGCTCAGTGGCTTCATATGAAATCTTTC	120
QY	1056	TGACATTGATCAAGATGAAAAAATCACTGTCAGAGAAATTTATCTAGCTATGACCTAAT	1115
Db	121	TGACATTGATCAAGATGAAAAAATCACTGTCAGAGAAATTTATCTAGCTATGACCTAAT	179
QY	1116	TGATGTTGCCATGTCGTGTCACGCACTGCGCCGCTCTGCTTCAGAAATACATCCTCC	1175
Db	180	TGATGTTGCCATGTCGTGTCACGCACTGCGCCGCTCTGCTTCAGAAATACATCCTCC	239
QY	1176	TTCCTTCGAGAGAGTGGCTCCGGCAGTGGGATTCGGTCAATAGCTCTTCTTGGA	1235
Db	240	TTCCTTCGAGAGAGTGGCTCCGGCAGTGGGATTCGGTCAATAGCTCTTCTTGGA	299
QY	1236	TCAAGAGCTGCTTAGAGAGCCGTCGTGTCAGAGATGAGCAGCAGCAGAGAAATCGCC	1295

```

|||||
Db 300 TCAGAGGCTGCTCAGAGAGCCCTCCTCAGAGGATGAGCAGCAGCCAGAGAAACTGTC 359
|||
Qy 1296 TGTGACATTTGAATTAAGAGCGGGAGAACTTCGAGCGAGGAGCTGAGAGTGGAGAA 1355
|||
Db 360 TGTGACATTTGAATTAAGAGCGGGAGAACTTCGAGCGAGGAGTGTGAGCTGGAGAA 419
|||
Qy 1356 GCCCGCCCAAGCCCTCTTGGAGCAGCGCAAAAGAGCAGAGCGGTTGCTCAGCTGGA 1415
|||
Db 420 GCCCGCCCAAGCCCTCTTGGAGCAGCGCAAAAGAGCAGAGCGGTTGCTCAGCTGGA 479
|||
Qy 1416 GCCCGCCGAGCAGAGAGAGAGAGCGGAGCGCCGAGAGCAGAGCGCAAGCGGAGCT 1475
|||
Db 480 GCCCGCCGAGCAGAGAGAGAGAGCGGAGCGCCGAGAGCAGAGCGCAAGCGGAGCT 539
|||
Qy 1476 GAGCTGTGAGAACACCTGTGAGAGAGCAGCGGAGCTGAGCGGAGCAGAGAGAGAG 1535
|||
Db 540 GAGCTGTGAGAACACCTGTGAGAGAGCAGCGGAGCTGAGCGGAGCAGAGAGAGAG 599
|||
Qy 1536 GAGAGAGAGATGAGAGAGCGCGAGCGCCGCAAAACCGGAACCTGGAAGCGAGCACT 1595
|||
Db 600 GAGAGAGAGATGAGAGAGCGCGAGCGCCGCAAAACCGGAACCTGGAAGCGAGCACT 659
|||
Qy 1596 TGAATGGAA-C-GGAACCGAGACAGGAACCTCTGAAATCAGAGAAACAGAGCAGAG 1654
|||
Db 660 TGAATGGAA-C-GGAACCGAGACAGGAACCTCTGAAATCAGAGAAACAGAGCAGAG 719
|||
Qy 1655 GC--ACCTGTGCTTGAAGCAGAGAGAGAGACTGTGAGATTGAGTTGAACCTGAA 1712
|||
Db 720 GCAACCGGCTGCTTGAAGCAGAGAGAGAGACTGTGAGATTGAGTTGAACCTGAA 779
|||
Qy 1713 TGACAAAAGCATCAGCTAGTAAGAGAAACTTCAGGATATCAGT-GTCGATGGCAACC 1771
|||
Db 780 TGACAAAAGCATCAGCTAGTAAGAGAAACTTCGAGATATCAGTGGTCCACTGGCAACC 839
|||
Qy 1772 AGA-GGCAAGAAATTGAGAGCAGC-AAAGCTTGA-GAGCTTAAGATTGC-TGAATC 1827
|||
Db 840 AGAGGCGCAAGAAATTGAGAGCAGCAGAAACAAGTCTAAAGAGCTAAAAATTGCTTGAATC 899
|||
Qy 1828 ACCCACTT 1835
|||
Db 900 ACCCACTT 907
|||

```

**RESULT 4**  
**BM950476**  
**LOCUS** BM950476 732 bp mRNA linear EST 14-MAR-2002  
**DEFINITION** UI-M-EHOP-bun-o-05-0-UI.r1 NIH\_BMAP\_EHOP Mus musculus cDNA clone  
**ACCESSION** IMAGE:5687260 5', mRNA sequence.  
**VERSION** BM950476.1 GI:19434066  
**KEYWORDS** EST.  
**SOURCE** house mouse.  
**ORGANISM** Mus musculus.

**REFERENCE**  
**AUTHORS** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NIH-MGC http://mgc.ncl.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
**TITLE** Unpublished (1999)  
**JOURNAL** Contact: Robert Strausberg, Ph.D.  
**COMMENT** Email: cgabs-r@mail.nih.gov  
 Tissue Procurement: Dr. James Lin, University of Iowa  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LINL at:  
 http://image.llnl.gov  
 This clone was contributed by the Brain Molecular Anatomy Project  
 (BMAP)  
 Seq primer: pYX-5.  
 Location/Qualifiers

```

source
1. 732
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5687260"
/clone_1db="NIH BMAP EHOP"
/tissue_type="Whole Brain"
/dev_stage="embryo 18.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/notes="Organ: brain; Vector: pYX-Asc; Site: 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldi, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I, and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is CAGCCAGAC. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."
BASE COUNT 208 a 172 c 201 g 148 t 3 others
ORIGIN
Query Match 14.2%; Score 722.8; DB 14; Length 732;
Best Local Similarity 99.3%; Pred. No. 1.1e-161;
Matches 724; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 3075 AGACATGTGTGTGTTGGAGAACTTCAAGTCAAGAGGTTGTTCCCAAGCTTTACT 3134
Db 4 ACATNNGTGTGTGTTGGAGAACTTCAAGTCAAGAGGTTGTTCCCAAGCTTTACT 63
Qy 3135 GAACTCAATTCAGGCGCCGTAAGGAATCCACAAGCATGATCTGGCCCTACTGAAAG 3194
Db 64 GAACTCAATTCAGGCGCCGTAAGGAATCCACAAGCATGATCTGGCCCTACTGAAAG 123
Qy 3195 TCCTCTAGTCTTAAGAGAGTGTCTCCCGCGCCCAAGCCAGCATTTCCGAGAAAG 3254
Db 124 TCCTCTAGTCTTAAGAGAGTGTCTCCCGCGCCCAAGCCAGCATTTCCGAGAAAG 183
Qy 3255 GTTTATGTCATGTACACATACAGAGTCTTGAGCAGAGAGATTAACTTTACGCAAG 3314
Db 184 GTTTATGTCATGTACACATACAGAGTCTTGAGCAGAGAGATTAACTTTACGCAAG 243
Qy 3315 GGATGTGATTGTGTACCAAGAAAGATGTGACTGTGAGCGGGAAGTGTGGGCAAA 3374
Db 244 GGATGTGATTGTGTACCAAGAAAGATGTGACTGTGAGCGGGAAGTGTGGGCAAA 303
Qy 3375 GTCCGAGTCTTCCCTTCTAATATGTAGAGGCTTAAAGTTCAAGAGGCTGTGAATGC 3434
Db 304 GTCCGAGTCTTCCCTTCTAATATGTAGAGGCTTAAAGTTCAAGAGGCTGTGAATGC 363
Qy 3435 TGGGAAAACAGGAGGTTTGAAGAAAAAAGCTGAAATTCGCCAGTTATTCTTACGC 3494
Db 364 TGGGAAAACAGGAGGTTTGAAGAAAAAAGCTGAAATTCGCCAGTTATTCTTACGC 423
Qy 3495 TGTACTGTGTCCCGAAACAATCACCCTGCTCTGGGAGCTGTGATCCGAAAAA 3554
Db 424 TGTACTGTGTCCCGAAACAATCACCCTGCTCTGGGAGCTGTGATCCGAAAAA 483
Qy 3555 GAAACCCAGTGTGTGGAAGAGAACTGCAGAGTGTGAGGAAAAAGCGCCAGATAGG 3614
Db 484 GAAACCCAGTGTGTGGAAGAGAACTGCAGAGTGTGAGGAAAAAGCGCCAGATAGG 543
Qy 3615 GTGTTTTCAGCAATTTATGTAAACTTTAAGCCCGGAGACAAATATACCCCAAC 3674
Db 544 GTGTTTTCAGCAATTTATGTAAACTTTAAGCCCGGAGACAAATATACCCCAAC 603
Qy 3675 TGAGTACCCAGACCGGAGTGTGAGCAGAGTGTGAGGATGTACGATTA 3734

```

Db 604 TGAGCTACCCAGACCGCACTGCAAGCCAGAGTGTGCAAGTATCGGATGTACATTA 663  
Qy 3735 CACCGCCAGACGATGACCACTAGCTTACAGCAAGGCGAGATCATCAAGCTCTCAA 3794  
Db 664 CACCGCCAGACGATGACCACTAGCTTACAGCAAGGCGAGATCATCAAGCTCTCAA 723  
Qy 3795 CAAGAGAGA 3803  
Db 724 CAAGAGAGA 732

RESULT 5  
LOCUS BM944544 738 bp mRNA linear EST 14-MAR-2002  
DEFINITION UI-M-EHOP-bvr-j-08-0-UI.r1 NIH\_BMAP\_EHOP Mus musculus cDNA clone  
IMAGE:5695975 5', mRNA sequence.  
BM944544  
VERSION BM944544.1 GI:19428129  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 738)  
TITLE NIH-MGC http://mgi.nci.nih.gov/.  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-r@mail.nih.gov  
Tissue Procurement: Dr. James Lin, University of Iowa  
cDNA library preparation: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
This clone was contributed by the Brain Molecular Anatomy Project  
(BMAP)

FEATURES  
source  
Location/Qualifiers  
1..738  
/organism="Mus musculus"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5695975"  
/clone\_lib="NIH\_BMAP\_EHOP"  
/tissue\_type="Whole Brain"  
/dev\_stage="embryo 18.5 dpc"  
/lab\_host="DH10B (T1 phage resistant)"  
/note="Organ: brain; Vector: pYX-ASC; Site 1: Ecor I;  
Site 2: Not I; The library was constructed according to  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
1996. Denatured mRNA was size fractionated on a 1% agarose  
gel. First strand cDNA synthesis was primed with an  
oligo-dT primer containing a Not I site. Double stranded  
cDNA was size selected according to mRNA size fraction,  
ligated with Ecor I adaptor, digested with Not I, and then  
cloned directionally into pYX-ASC vector. The library tag  
sequence located between the Not I site and the polyA tail  
is CAGCCACGAC. This library was created for the  
University of Iowa Mouse Brain Molecular Anatomy Project  
(BMAP): 'Gene Discovery in the Developing Mouse Nervous  
System', supported by National Institute of Mental Health  
(NIMH), Hemm Chin, Ph.D., program coordinator."

BASE COUNT 198 a 213 c 184 g 143 t  
ORIGIN  
Query Match 14.2%; Score 721.2; DB 14; Length 738;  
Best Local Similarity 99.5%; Pred. No. 2.5e-161;  
Matches 734; Conservative 0; Mismatches 3; Indels 1; Gaps 1;  
Qy 2568 CCAAGTGTGAGAGCCGATGCTTGGAGAGAGCTGAAGAGAGCGGATGTTCCC 2627

Db 1 CCAAGTGTGAGAGCCGATGCTTGGAGAGAGCTGAAGAGAGCGGATGTTCCC 60  
Qy 2628 TGCAAACTATGACAGAAAAGATTTCCAGAAAATGAGTTCCCACTCCAGCCAAACAGTAGAC 2687  
Db 61 TGCAAACTATGACAGAAAAGATTTCCAGAAAATGAGTTCCCACTCCAGCCAAACAGTAGAC 120  
Qy 2688 CGATCTGACATCTGCCCCCTGCCCCCAACTGCTCTGCTGAGACCCCTGCTCTTTGCC 2747  
Db 121 CGATCTGACATCTGCCCCCTGCCCCCAACTGCTCTGCTGAGACCCCTGCTCTTTGCC 180  
Qy 2748 AGTACCTCTCTGAGCCCTCCCAACCCCAACCACTGGGCGAGCTTCACTCCAGCTG 2807  
Db 181 AGTACCTCTCTGAGCCCTCCCAACCCCAACCACTGGGCGAGCTTCACTCCAGCTG 240  
Qy 2808 GCCCAGAGCTCAACAGAGAGAGAGCCAGAGAGCACTGGGATATCGTGGCGGCTAGCC 2867  
Db 241 GCCCAGAGAGCTCAACAGAGAGAGAGCCAGAGAGCACTGGGATATCGTGGCGGCTAGCC 300  
Qy 2868 TTCTCTGACGCTACCTTGTGTGTGCTGCGCAGTTACGGCAGAGATCAGCTTTACCCAGCCAC 2927  
Db 301 TTCTCTGACGCTACCTTGTGTGTGCTGCGCAGTTACGGCAGAGATCAGCTTTACCCAGCCAC 360  
Qy 2928 AGCAGCTGGCTCTCCCTCCCATCTCCCGTCTGGGCGAGGAGAGAGGTTGAGAGGCTACA 2987  
Db 361 AGCAGCTGGCTCTCCCTCCCATCTCCCGTCTGGGCGAGGAGAGGTTGAGAGGCTACA 420  
Qy 2988 AGCCAGAGCCCTGTATCTCTGAGAGCCAAAGAGCAACACTTAATTTTAAACAAAG 3047  
Db 421 AGCCAGAGCCCTGTATCTCTGAGAGCCAAAGAGCAACACTTAATTTTAAACAAAG 480  
Qy 3048 TGAGGTATATACGCTTGTGAGACAGAGCATGTGTGTTGGAGAGTTCAAGGTCA 3107  
Db 481 TGAGGTATATACGCTTGTGAGACAGAGCATGTGTGTTGGAGAGTTCAAGGTCA 540  
Qy 3108 GAAGGTTGTTTCCCAAGCTTACGTAAGAACTATTTCAGGAGCCGTAAGAAATCCAC 3167  
Db 541 GAAGGTTGTTTCCCAAGCTTACGTAAGAACTATTTCAGGAGCCGTAAGAAATCCAC 600  
Qy 3168 AAGCATGATATCTGCGCTTACTGAAAGTCTGCTAGTCTTAAAGAGTGGCTTCCCCGGC 3227  
Db 601 AAGCATGATATCTGCGCTTACTGAAAGTCTGCTAGTCTTAAAGAGTGGCTTCCCCGGC 660  
Qy 3228 CGCCAGCCAGCCATCTCCCGAGAGAGATTATGTCATATACATACAGAGATCTGCA 3287  
Db 661 CGCCAGCCAGCCATCTCCCGAGAGAGATTATGTCATATACATACAGAGATCTGCA 719  
Qy 3288 GCAAGAGATTAACTT 3305  
Db 720 GCAAGAGATTAACTT 737

RESULT 6  
LOCUS BG922629 782 bp mRNA linear EST 05-JUN-2001  
DEFINITION 60282158F1 NCI\_CGAP\_Mame Mus musculus cDNA clone IMAGE:4950524 5',  
mRNA sequence.  
BG922629  
ACCESSION BG922629 GI:14303105  
VERSION BG922629.1  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 782)  
TITLE NIH-MGC http://mgi.nci.nih.gov/.  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-r@mail.nih.gov  
Tissue Procurement: Jeffrey Green M.D.  
cDNA library preparation: Life Technologies, Inc.  
DNA Sequencing by: Inocyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov  
plate: LLNL0905 row: e column: 21  
High quality sequence stop: 716.

## FEATURES

source

1..782

Location/Qualifiers

/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4950524"  
/clone\_1b="NCI\_CGAP\_Mam6"  
/sex="female, virgin"  
/tissue\_type="infiltrating ductal carcinoma"  
/dev\_stage="5 months"  
/lab\_host="DH10B"

/note="Organ: mammary; Vector: pCMV-Sport6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT

224 a 187 c 191 g 180 t

ORIGIN

Query Match 13.6%; Score 691.2; DB 13; Length 782;  
Best Local Similarity 97.4%; Pred. No. 3.8e-154;  
Matches 746; Conservative 0; Mismatches 13; Indels 7; Gaps 4;

3567 ATGCTGGAGAGAGAACTGACAGCTGAGGAGAAAGCCCAATAGGCTGTTCCAGC 3626  
1 ATGCTGGAGAGAGAACTGACAGCTGAGGAGAAAGCCCAATAGGCTGTTCCAGC 60  
3627 -AAATATGCAACTCTTAAGCCCGGAGCAAGCAAAATACCCCACTGAGCTACCA 3685  
61 GAAATATGCAACTCTTAAGCCCGGAGCAAGCAAAATACCCCACTGAGCTACCA 120  
3686 AGACCCGAGTGCAGCCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 3745  
121 AGACCCGAGTGCAGCCGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 180  
3746 AGATGAGCAACTGAGCTTCAAGCAAGCCCAATAGGCTGTTCCAGC 3805  
181 AGATGAGCAACTGAGCTTCAAGCAAGCCCAATAGGCTGTTCCAGC 240  
3806 CGGACTGTGAGAGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 3865  
241 CGGACTGTGAGAGAGAGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 300  
3866 AGCTGACCAAGACATGAGCCCGAGCAGCAATGATGTCATGTCATGTCATGTCATG 3925  
301 AGCTGACCAAGACATGAGCCCGAGCAGCAATGATGTCATGTCATGTCATGTCATG 360  
3926 AGCTGAGAAAGTCTTCAAGAGCCCAATGTCATGTCATGTCATGTCATGTCATGTCAT 3985  
361 AGCTGAGAAAGTCTTCAAGAGCCCAATGTCATGTCATGTCATGTCATGTCATGTCAT 420  
3986 AGATGAGCAACTGAGCTTCAAGCAAGCCCAATAGGCTGTTCCAGC 4445  
421 AGATGAGCAACTGAGCTTCAAGCAAGCCCAATAGGCTGTTCCAGC 480  
4446 AGCTGACCAAGACATGAGCCCGAGCAGCAATGATGTCATGTCATGTCATGTCATG 4105  
481 AGCTGACCAAGACATGAGCCCGAGCAGCAATGATGTCATGTCATGTCATGTCATG 540  
4106 TCACCTGCGTGCAGAGGAGAAATGCAAGCTGCAAGGGTGTGGTCTTCTTGG 4165  
541 TCACCTGCGTGCAGAGGAGAAATGCAAGCTGCAAGGGTGTGGTCTTCTTGG 600  
4166 GCGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 4225  
601 GCGTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 657  
4226 TAAGATTATTTTAAATGATTTAGCTTTTAAATTAATCAATCAATCAATCAATCAAT 4285

Db 658 TAAATATTTTAAATGATTTTAAATTAATCAATCAATCAATCAATCAATCAAT 716  
4286 TTGCTATTTTGGTTTAC-AAAAACACCACATCAAGAGTGC 4329  
717 TTGCTATTTTGGTTTACAAAAAAGACACATCAAGAGTGC 762

## RESULT 7

BO941336

LOCUS

DEFINITION

5, mRNA sequence.

ACCESSION

BO941336

VERSION

BO941336.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 896)

NIH-MGC http://mgs.nci.nih.gov/.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgabs-remail.nih.gov

Tissue Procurement: DCTD/DTP/Gazdar

CDNA Library Preparation: Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LCM2594 row: K column: 01

High quality sequence stop: 763.

## FEATURES

source

1..896

Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6420600"  
/clone\_1b="NIH\_MGC\_18"  
/tissue\_type="large cell carcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: lung; Vector: pGEMT; Site\_1: XhoI; Site\_2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(C). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC Library."

BASE COUNT

214 a 232 c 230 g 216 t 4 others

ORIGIN

Query Match 13.0%; Score 660.2; DB 14; Length 896;  
Best Local Similarity 86.8%; Pred. No. 9.9e-147;  
Matches 787; Conservative 0; Mismatches 105; Indels 15; Gaps 5;

40 GCTTGTTGCTCCGTAAGTACGCGCGCTGCAAGAGAGATCCCAAGCGGCTCCGGAGC 99  
1 GCTTGTTGCTCCGTAAGTACGCGCGCTGCAAGAGAGATCCCAAGCGGCTCCGGAGC 60  
100 GCGGGAGGACAGAGGCGGCGGCGGAGTGTGTGCGGCGTCCGAGCTCCGCGCTTC 159  
61 -----GACAGAGAGGCGGCGGCGGAGTGTGTGCGGCGTCCGAGCTCCGCGCTTC 109  
160 CTCGC-GCGGCGTGGCGGCTGCACTGATTTGTGAGGGGCGGCGCGCAAGCGCGCG 218  
110 CTCGACGCGGCGGCGGAGGCGGAGCTGATTTGTCTCGGCGGCGGAGCGCGGCGCG 169  
219 GAGATGAGGCGTGCATCAGCAAGGTGAACGTAATGAACATGCTGATTTCCACAGC 278  
170 GAGATGAGGCGTGCATCAGCAAGGTGAACGTAATGAACATGCTGATTTCCACAGC 229

QY 279 TTTCGGTGTAGCTGAGATCTGGGCAATACGTGAGAGAAAGGCCAAGCATGACCA 338  
 DB 230 TTTTGGTGGAGCGCTGGATATCTGGGCCAATACGTAGAGAAAGCGAAGATGATCA 289  
 QY 339 GCAGTTCCTTAGCTGAGAGCGATAGCGGGATTTATTCGCGGATCAAGCAGGAACT 398  
 DB 290 GCAGTTCCTTAGCTGAGAGCGATAGCGGGATTTATTCGCGGATCAAGCAGGAACT 349  
 QY 399 TTTTTCCTAGCTGGGTATCTAGCGCTGCTTTAGCACAATATGGCGCTAGCGACAT 458  
 DB 350 TTTTTCCTAGCTGGGTATCTAGCGCTGCTTTAGCACAATATGGCGCTAGCGACAT 409  
 QY 459 GAATACGATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 518  
 DB 410 GAATACGATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 469  
 QY 519 GAAGTACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 578  
 DB 470 GAAGTACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 529  
 QY 579 GGTATTTTCCAGTGCACACGATTTGGTATAGAGAGATTTCTAGCATGCCACTCAC 638  
 DB 530 TGTCTATTTTACGCGACACGATTTGGTATAGAGAGATTTCTAGCATGCCACTCAC 589  
 QY 639 AGCTGTGCTCCTGCGCAATGGGCTCCATTCAGTGTGATGATGATGATGATGATGAT 698  
 DB 590 AGCTGTGCTCCTGCGCAATGGGCTCCATTCAGTGTGATGATGATGATGATGATGAT 649  
 QY 699 ATCTTGTCTCCTGCGCAATGGGCTCCATTCAGTGTGATGATGATGATGATGATGAT 758  
 DB 650 ATCTTGTCTCCTGCGCAATGGGCTCCATTCAGTGTGATGATGATGATGATGATGAT 709  
 QY 759 GCTCTGCTGCTGCTTGGCGATCTTGCAGCCATGAGGCGCAAGAGTCTTCC-TTCAGA 817  
 DB 710 ACCTGTGCTGCTGCTTGGCGATCTTGCAGCCATGAGGCGCAAGAGTCTTCC-TTCAGA 769  
 QY 818 GATCTGTGCTGCTGCTTGGCGATCTTGCAGCCATGAGGCGCAAGAGTCTTCC-TTCAGA 877  
 DB 770 GATCTGTGCTGCTGCTTGGCGATCTTGCAGCCATGAGGCGCAAGAGTCTTCC-TTCAGA 829  
 QY 878 CCAGCG-CCCCCTCAGCAGCAGATGG-GCTGTGCTCAGTCAATCAAGGCTAAATACAG 935  
 DB 830 CCAGTGTCTCAGCAGCAGATGG-GCTGTGCTCAGTCAATCAAGGCTAAATACAG 889  
 QY 936 GCAGTTA 942  
 DB 890 GCATTTA 896

RESULT 8  
 LOCUS BQ719508 979 bp mRNA linear EST 16-JUL-2002  
 DEFINITION AGENCOURT 8219793 lupski\_sympathetic\_trunk Homo sapiens cDNA clone  
 IMAGE:6188111 5', mRNA sequence.  
 ACCESSION BQ719508  
 VERSION BQ719508.1 GI:21858405  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 979)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Dr. James R. Lupski  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:

http://image.lnl.gov  
 Plate: L1M13583 row: c column: 24  
 High quality sequence stop: 643.  
 Location/Qualifiers  
 1..979  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="IMAGE:6188111"  
 /issue\_type="sympathetic trunk"  
 /dev stage="adult", 16 yr  
 /lab\_host="DH10B"  
 /note="Vector: PCW-SPORE (Life Technologies); Site\_1:  
 Note: Site 2: SalI; cDNA made by oligo-dT priming.  
 Directionally cloned using the following adaptors:  
 5'-TCGACCCAGCGCG-3' and  
 5'-GACTAGTCTAGATGCGAGCGCGCCCT(15)-3'. Site selected >  
 1 kb for average insert length 1.9 kb. This is a primary  
 library, non-amplified. Library constructed by Life  
 Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor  
 College of Medicine); available through Life  
 Technologies."  
 BASE COUNT 294 a 247 c 218 g 217 t 3 others  
 ORIGIN  
 Query Match 12.9%; Score 658.2; DB 14; Length 979;  
 Best Local Similarity 84.2%; Pred. No. 3e-146;  
 Matches 776; Conservative 0; Mismatches 141; Indels 5; Gaps 3;  
 QY 549 ACTTCCCCCTGATGAAACAGACAGAGGCTATTTCCAGTGCACACGATTTGGTAT 608  
 DB 1 ACTTCCCCCTGATGAAACAGACAGAGGCTATTTCCAGTGCACACGATTTGGTAT 60  
 QY 609 AGAGAGATTTCTAGATGCGACACACTCAGAGCTGTGCTCTCTGTCCTGTCAT 668  
 DB 61 GGGAGATATGCGCAGATGCGACCGCTTACAGCTGTGCTCCAGTGCATGGGATCCAT 120  
 QY 669 TCCAGTGTGGAATGTCACACCTTATGATCTTCTGTCCTCCAGCAGAGCTGCTCC 728  
 DB 121 TCCAGTGTGGAATGTCACACCTTATGATCTTCTGTCCTCCAGCAGAGCTGCTCC 180  
 QY 729 CTTGGCTAAGGGGCTCTCCCTCATACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCT 788  
 DB 181 CTTGGCTAAGGGGCTCTCCCTGTATACACCTCTGCTGCTGCTGCTGCTGCTGCTGCT 240  
 QY 789 CACATGGCCAAAGAGTCTTCTTCAAGAGATCTGCTCAAGGCTCAATTAACCTAA 848  
 DB 241 CACATGGCCAAAGAGTCTTCTTGTATGATCTGATGATGATGATGATGATGATGAT 300  
 QY 849 GTTACAGAGGCAATCATTTGATGCGCCAGCGCCCTCCAGCAGAGATGGGCTGT 908  
 DB 301 ATTACAAAGGCAAGTCAATTTATGATGCGCAAGTGTCCACAGAGAGAGGCTGT 360  
 QY 909 GCTTCACTCAATCAAGGCTGAATATAGGAGGATTTTCAACAGCAGCAAACTATGAG 968  
 DB 361 TCTCACTCAATCAAGGCTGAATATAGGAGGATTTTCAATAGCATGACAACTATGAG 420  
 QY 969 TGGACCTTAAAGGTCCTCCAGCAGAGATATTTCTCAAGGATGAAATGCTGCTGCT 1028  
 DB 421 TGGACCTTAAAGGTCCTCCAGCAGAGATATTTCTTATGATGATGATGATGATGAT 480  
 QY 1029 TCAGCTGGCTTCAATATGAGATCTTCTGATGATGATGATGATGATGATGATGATGAT 1088  
 DB 481 TCAGCTGGCTTCAATATGAGATCTTCTGATGATGATGATGATGATGATGATGATGAT 540  
 QY 1089 AGAATTTATCTTACATGACCTAATGATGATGATGATGATGATGATGATGATGAT 1148  
 DB 541 GGAATTTATCTTACATGACCTAATGATGATGATGATGATGATGATGATGATGAT 600  
 QY 1149 CGTCTGCTCCAGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1208  
 DB 601 TGTCTGCTCCAGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660





TITLE  
JOURNAL  
COMMENT

M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takega, Y., Tanaka, T., Toya, T., Muramatsu, M., and Hayashizaki, Y.  
RIKEN Mouse ESTs (Arikawa, T., et al. 2001)  
Unpublished (2001)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.go.jp  
URL: http://genome.gsc.riken.go.jp/  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. *Genome Res.* 10 (10), 1617-1630 (2000)  
wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)  
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y., and Hayashizaki, Y.  
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 261-269 (2001)  
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K., and Hayashizaki, Y.  
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. *Mamm. Genome* 12, 673-677 (2001)  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.  
e mouse tissues.

FEATURES  
source

Location/Qualifiers  
1. 650  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="D130043B15"  
/clone\_1ib="RIKEN full-length enriched, 12 days embryo spinal ganglion"  
/tissue type="spinal ganglion"  
/dev stage="12 days embryo"  
/lab\_host="DH10B"  
/note="Site 1: Salt; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 15'  
GAGAGAGAGAGCGCGCGCACTCGACTTTTCTTTTNN 3', cDNA was prepared by using trehalose chemo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence 15'  
GAGAGAGAGATTCGAGTTAATTAATTAATCCCGCCCCCCC 3'. cDNA was cleaved with BamHI and XhoI. Vector: a modified plasmid vector KS(+) after bulk excision from Lambda FLC I."

BASE COUNT  
ORIGIN

Query Match 12.5%; Score 633.6; DB 10; Length 650;  
Best local Similarity 99.2%; Pred. No. 2,2e-140;  
Matches 636; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2463 CGAGAGTGTAAAGTGTATATATACCGAGCGCTGTACCCCTTTGATTCAGAAATCGCA 2522  
Db 1 GGAGAGTGTAAAGTGTATATATACCGAGCGCTGTACCCCTTTGATTCAGAAATCGCA 60  
Qy 2523 TGAATATCCATTCAGCCAGAGATATATGTCATGTGTGATGAAGCCAGCTGAGAGCC 2582  
Db 61 TGAATATCCATTCAGCCAGAGATATATGTCATGTGTGATGAAGCCAGCTGAGAGCC 120  
Qy 2583 AGATATGCTTGAAGAGAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2642  
Db 121 AGATATGCTTGAAGAGAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180  
Qy 2643 AAAGATTCAGAAATATAGTGTCCCATCTCCAGCCAAACGCTGACGATCTGACATTCG 2702  
Db 181 AAAGATTCAGAAATATAGTGTCCCATCTCCAGCCAAACGCTGACGATCTGACATTCG 240  
Qy 2703 CCGTGTCCCAACCTGCTGTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2762  
Db 241 CCGTGTCCCAACCTGCTGTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300  
Qy 2763 GCGCTTCACAAACCCCAACCACTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2822  
Db 301 GCGCTTCACAAACCCCAACCACTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360  
Qy 2823 CGAGAGCCAGAAACGAGCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2882  
Db 361 CGAGAGCCAGAAACGAGCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420  
Qy 2883 TAGTGTCTGCGAGTTAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2942  
Db 421 TAGTGTCTGCGAGTTAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480  
Qy 2943 CCCATCTCCGCTCTGCGCCAGAGGTGAAGAGGTGAAGAGGTGAAGAGGTGAAGAGGTGA 3002  
Db 481 CCCATCTCCGCTCTGCGCCAGAGGTGAAGAGGTGAAGAGGTGAAGAGGTGAAGAGGTGA 540  
Qy 3003 TCCCTGAGAGACCAAAAGAGCAACCACTTAATTTAAACAAAGTACGCTCATCCGT 3062  
Db 541 TCCCTGAGAGACCAAAAGAGCAACCACTTAATTTAAACAAAGTACGCTCATCCGT 600  
Qy 3063 TCTGAAACAGACAGATGTGTGTGTTGGAGAAAGTTCAAG 3103  
Db 601 TCTGAAACAGACAGATGTGTGTGTTGGAGAAAGTTCAAG 641

RESULT 11  
B0898718 1062 bp mRNA linear EST 16-AUG-2002  
LOCUS  
DEFINITION  
AGENCOURT 8118759 Lupski dorsal root ganglion Homo sapiens cDNA  
clone IMAGE:6180303 5', mRNA sequence.  
B0898718  
VERSION  
B0898718.1 GI:22290732  
KEYWORDS  
EST.  
SOURCE  
ORGANISM  
human.  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1062)  
NIH-MGC <http://mgc.ncl.nih.gov/>.  
NATIONAL INSTITUTES OF HEALTH, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: rgs@ncl.nih.gov  
Tissue Procurement: Dr. James R. Lupski  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L1AM13562 row: n column: 16  
High quality sequence start: 116  
High quality sequence stop: 760.



## FEATURES

## source

Location/Qualifiers  
 1. 1062  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6180303"  
 /clone\_1ib="Lupski\_dorsal\_root\_ganglion"  
 /sex="male"  
 /tissue\_type="dorsal root ganglia"  
 /dev\_stage="adult, 36 yr"  
 /lab\_host="DH10B"  
 /note="vector: pCMV-SPORT6 (Life Technologies); Site\_1:  
 NotI; Site\_2: SalI; cDNA made by oligo-dT priming.  
 directionally cloned using the following adaptors:  
 5'-TCGACCCGACGCTCG-3' and  
 5'-TCGACGCTTACATGATCGGCGCGCCCT(15)-3'. Size selected >  
 1 kb for average insert length 1.7 kb. This is a primary  
 library, non-amplified. Library constructed by Life  
 Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor  
 College of Medicine) and is available through Life  
 Technologies."

BASE COUNT 338 a 250 c 261 g 209 t 4 others

## ORIGIN

Query Match 12.4%; Score 631.4; DB 14; Length 1062;  
 Best Local Similarity 90.9%; Pred. No. 7.8e-140;  
 Matches 715; Conservative 0; Mismatches 68; Indels 4; Gaps 4;

3242 TTCGCGGAGAGGTTTATTCGATGACATGAGAGTTCTGAGCAAGGATTAA 3301  
 113 TTGGCGGAGAGGTTTATTCGATGACATGAGAGTTCTGAGCAAGGATTAA 172  
 3302 CCTTTCAGCAAGGAGGATGATGTTGTTTACCAAGAAAGTGTGACGCGGAA 3361  
 173 CCTTTCAGCAAGGAGGATGATGTTGTTTACCAAGAAAGTGTGACGCGGAA 232  
 3362 CGGTGGGCGCAAGTCCGAGTCTTCTCTAATGAGAGCTTTAAAGATTCAAG 3421  
 233 CAGTGGGCGCAAGTCCGAGTCTTCTCTAATGAGAGCTTTAAAGATTCAAG 292  
 3422 GCTCTGGAATGCTGGGAAACAGGAGTTAGAAAAAACTGAATTTGCCAGTTA 3481  
 293 GCTCTGGAATGCTGGGAAACAGGAGTTAGAAAAAACTGAATTTGCCAGTTA 352  
 3482 TTGCTTCTACGCTGTACTGTATGCTCCGACCAACTCACTGCTGCGACGTTAT 3541  
 353 TTGCTTCTACGCTGTACTGTATGCTCCGACCAACTCACTGCTGCGACGTTAT 412  
 3542 TGATCCGAAAAAGAACCCAGGTGATGTTGGAGAGAGTTCGAGCGGAAAA 3601  
 413 TGATCCGAAAAAGAACCCAGGTGATGTTGGAGAGAGTTCGAGCGGAAAA 472  
 3602 AGGCGCGAGTAGGGTGGTTCCAGCAATATGCTTAAGCCCGGAAACAGCA 3661  
 473 AGGCGCGAGTAGGGTGGTTCCAGCAATATGCTTAAGCCCGGAAACAGCA 532  
 3662 AAATCACCCCACTGAGCTACCCCAAGCCGAGTGCAGCAGAGTGTGCGAGTAT 3721  
 533 AAATCACCTCCCAAGCAGCCTTAAGTCAACAGCATTAGCGGCGAGTGTGCGAGTAT 592  
 3722 GGATGTACGATTACCGCCCAAGACATGACCACTTACGCTTCAAGAAAGGCAATCA 3781  
 593 GGATGTACGATTACCGCCCAAGACATGACCACTTACGCTTCAAGAAAGGCAATCA 652  
 3782 TCAACGTCCTCAACAGAGAGACCCGAGTGTGGAAGAGAAAGTCAAGGCAAGTTG 3841  
 653 TCAACGTCCTCAACAGAGAGACCCGAGTGTGGAAGAGAAAGTCAAGGCAAGTTG 712  
 3842 GGCTCTTCCATCCCAATATGTTAAAGTGCACACAGCATGAGCCCAAGCAGCAATGAA 3901  
 713 GGCTCTTCCATCCCAATATGTTAAAGTGCACACAGCATGAGCCCAAGCAGCAATGAA 772  
 3902 TCATATGTTGTCATCCCGCTTCAAGCTTCTCAAGAGCCCACTATCCCAT 3961

Db 773 TCATATGTTGTCATCCCGCTTCAAGCTTGAAGTCTCAAGAGCCCACTATCCCAT 832

Qy 3962 ATCACTGCGCC-AGAGGAGATGATGG-AGATGACGCC-TTGATCATGTGA-CTTGACGACAT 4017

Db 833 ATCACTGCGCCAAAGGATGATGGAGATGACGCCCTTGATCATGTGACCTTCCAGCAT 892

Qy 4018 GATCACC 4024

Db 893 GATCACC 899

## RESULT 12

## AV246405

## DEFINITION

AV246405 RIKEN full-length enriched, 0 day neonate head Mus  
 musculus cDNA clone 4832408K23 3' similar to Af064243 Homo sapiens  
 interseccion short form mRNA, mRNA sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

AV246405 674 bp mRNA linear EST 24-OCT-2001  
 AV246405 RIKEN full-length enriched, 0 day neonate head Mus  
 musculus cDNA clone 4832408K23 3' similar to Af064243 Homo sapiens  
 interseccion short form mRNA, mRNA sequence.  
 AV246405.2 GI:16387096  
 EST.  
 house mouse.  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 674)  
 Hiramoto, K., Horii, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda  
 M., Koyu, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M.,  
 Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki  
 D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,  
 Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T.,  
 Muramatsu, M. and Hayashizaki, Y.  
 RIKEN Mouse ESTs (Arakawa, T., et al. 2001)  
 Unpublished (2001)  
 On Nov 4, 1999 this sequence version replaced gi:6233864.  
 Contact: Yoshihide Hayashizaki  
 Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@gsc.riken.go.jp,  
 URL: http://genome.gsc.riken.go.jp/  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh  
 M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
 Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura  
 S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and  
 Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)  
 Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara  
 Y. and Hayashizaki, Y.  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Kondo, S., Shingawa, A., Saito, T., Kiyosawa, H., Yamana, I., Aizawa  
 K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and  
 Hayashizaki, Y.  
 Computational Analysis of Full-length Mouse cDNAs Compared with  
 Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
 Please visit our web site (http://genome.gsc.riken.go.jp/) for  
 further details.  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in Riken  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.



Db	1	AGGAATTCACAGCATCGATTCTTGCCCTTACTGAAATCCTCGTAGCTTAAAGAGTGC	60
QY	3217	GCTTTCCTCCGCGCCGCAAGCCAGCCATTCCCGAGAAAGATTATTGCCATGTACACATAC	3276
Db	61	GCTTTCCTCCGCGCCGCAAGCCAGCCATTCCCGAGAAAGATTATTGCCATGTACACATAC	120
QY	3277	GAGAGTCTGAGCAAGAGATTTAACCTTTAGCAAGGGGATGTGATTGTGGTTACCAAG	3336
Db	121	GAGAGTCTGAGCAAGAGATTTAACCTTTAGCAAGGGGATGTGATTGTGGTTACCAAG	180
QY	3337	AAAGATGTGACTGTGTGAGCGGAAACGGTGGCGCAAGTCCGGAGTCTTCCCTTTAC	3396
Db	181	AAAGATGTGACTGTGTGAGCGGAAACGGTGGCGCAAGTCCGGAGTCTTCCCTTTAC	240
QY	3397	TATGTGAGGCTTAAAGATTCAAGGGCTCTTGAACCTCTGGGAAAAACGGAGTTTACGA	3456
Db	241	TATGTGAGGCTTAAAGATTCAAGGGCTCTTGAACCTCTGGGAAAAACGGAGTTTACGA	300
QY	3457	AAAAAACTGMAATTGCCAGGTTATTCCTTCTTAACGCTGCTACGTGTCCTCCGAACAATC	3516
Db	301	AAAAAACTGMAATTGCCAGGTTATTCCTTCTTAACGCTGCTACGTGTCCTCCGAACAATC	360
QY	3517	ACCTGGCTCTCGGGCAGCTGATTTCTGATCCGGAAAAAAGAACCCAGGTGATGTGGGAA	3576
Db	361	ACCTGGCTCTCGGGCAGCTGATTTCTGATCCGGAAAAAAGAACCCAGGTGATGTGGGAA	420
QY	3577	GGAGAACTGCAAGCTCGAGGGGAAAAAGCCGAGATAGGGTGGTTTCAGCAATTATGTC	3636
Db	421	GGAGAACTGCAAGCTCGAGGGGAAAAAGCCGAGATAGGGTGGTTTCAGCAATTATGTC	480
QY	3637	AAACTTTTAAGCCCCCGGAACAACAATAATCAACCCACTGAGCTTACCAAGACCGCAAGTGC	3696
Db	481	AAACTTTTAAGCCCCCGGAACAACAATAATCAACCCACTGAGCTTACCAAGACCGCAAGTGC	540
QY	3697	CAGCAGAGAGTGTGCAGAGTGTATCGGGATGTATGATTCACCGCCAGACGATGACGA	3756
Db	541	CAGCAGAGAGTGTGCAGAGTGTATCGGGATGTATGATTCACCGCCAGACGATGACGA	600
QY	3757	CTAGCCTTCA 3766	
Db	601	CTAGCCTTCA 610	
RESULT 14			
LOCUS	A1594919	603 bp	linear EST 15-MAR-2000
DEFINITION	m1799g10.y1 Soares protein p3NMF19.5 Mus musculus cDNA clone IMAGE482370 5' similar to TR:042287 042287 INTERSECTIN.; mRNA sequence.		
ACCESSION	A1594919		
VERSION	A1594919.1	GI:4603967	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 603)		
AUTHORS	Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Giddons, M., Fape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.		
TITLE	The WashU-NCI Mouse EST Project 1999		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Marra M/WashU-NCI Mouse EST Project 1999 Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.		

This read is a RESEQUENCE of a previously sequenced mouse clone. This read has been VERIFIED (found to hit its original self in the correct orientation)  
 MGI:293114  
 Seq. primer: -40RP from Gibco  
 High quality sequence stop: 464  
 POLY-A=NO. *Annotation/Qualifiers*

FEATURES	Location/Qualifiers
source	1. .603

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:482370"
/clone_lib="Sceres mouse p3NMF19.5"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/notes="Vector: pU713D (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCACATCTCAAGTGGAGGCGCGCATTTTTTTTTTTTTTTT 3',
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pU713 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University)."

```

Query Match:	11.8%	Score 602;	DB 9;	Length 603;
Best Local Similarity	99.8%	Pred. No. 7,6e-133;		
Matches 602;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
QY 3392	GGAGATTAACTTTTCAGCAAGGGAGTGTGATTGTGTACCAAGAAAGTGTGACTGG	3351		
Db 1	GGAGATTAACTTTTCAGCAAGGGAGTGTGATTGTGTACCAAGAAAGTGTGACTGG	60		
QY 3352	TGCAACGGMAACGCTGGGCGCAACAAGTCGGAGCTCTCCCTCTMACTATGTGAGGCTTAA	3411		
Db 61	TGCAACGGMAACGCTGGGCGCAACAAGTCGGAGCTCTCCCTCTMACTATGTGAGGCTTAA	120		
QY 3412	GATTTCAGAGGGCTCTTGGAACCTCTGGGAAAAACAGGAGTTTATGGAAAAAAACCTGAAATT	3471		
Db 121	GATTTCAGAGGGCTCTTGGAACCTCTGGGAAAAACAGGAGTTTATGGAAAAAAACCTGAAATT	180		
QY 3472	GCCCAAGTTATTTGCTTCTTACGCTGCTACTGTCTCCCAACAATCTACCCGCTCTGGG	3531		
Db 181	GCCCAAGTTATTTGCTTCTTACGCTGCTACTGTCTCCCAACAATCTACCCGCTCTGGG	240		
QY 3532	CAGTGATTTGATTCGGAAAAAAGAACCCAGGTGATGTGGAAAGAGAACTGCAAGCT	3591		
Db 241	CAGTGATTTGATTCGGAAAAAAGAACCCAGGTGATGTGGAAAGAGAACTGCAAGCT	300		
QY 3592	CGAGGAAAAAAGCGCCAGATAGGGTGTGTTCCAGCAAAATTATGTCAAACTTTCTAAGCCCC	3651		
Db 301	CGAGGAAAAAAGCGCCAGATAGGGTGTGTTCCAGCAAAATTATGTCAAACTTTCTAAGCCCC	360		
QY 3652	GGAAACAAGCAAAATTCACCCCACTAGACTACCAAGACCGGAGTGCAGCCAGCAAGTGTG	3711		
Db 361	GGAAACAAGCAAAATTCACCCCACTAGACTACCAAGACCGGAGTGCAGCCAGCAAGTGTG	420		
QY 3712	CAGGTGATTCGGAGTGTACGATTACACCGCCAGAACGATGACGAATTAAGCTTCAGCAAA	3771		
Db 421	CAGGTGATTCGGAGTGTACGATTACACCGCCAGAACGATGACGAATTAAGCTTCAGCAAA	480		
QY 3772	GGCCAGATCATCAACGTCTCTCAACAAGAGAGACCGGACTGTGTGAAAAAGAGAAAGTACT	3831		
Db 481	GGCCAGATCATCAACGTCTCTCAACAAGAGAGACCGGACTGTGTGAAAAAGAGAAAGTACT	540		
QY 3892	GGGCAAGTGTGGGCTCTTCCCATCAATTTATGTAAGCTGACCAAGACATGAGACCCAGC	3891		
Db 541	GGGCAAGTGTGGGCTCTTCCCATCAATTTATGTAAGCTGACCAAGACATGAGACCCAGC	600		

QY 3892 CAG 3894  
Db 601 CAG 603

## RESULT 15

LOCUS AW910367 606 bp mRNA linear EST 25-MAY-2000  
DEFINITION ur80908.y1 NCI CGAP Mam6 Mus musculus cDNA clone IMAGE:3156638 5' similar to TR:Q920R4 Q920R4 EMBL PROTEIN. ; mRNA sequence.  
ACCESSION AW910367  
VERSION AW910367.1 GI:8075608  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 606)  
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
Tumor Gene Index  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Unpublished (1997)  
JOURNAL  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Jeffrey Green M.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone Distribution: NCI-CCAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
image.llnl.gov/image/html/resources.shtml

MG1:1059394  
Seq primer: -40RP from Gibco  
High quality sequence stop: 442.  
Location/Qualifiers  
1..606

FEATURES  
source  
/organism="Mus musculus"  
/strain="FVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3156638"  
/clone\_1ib="NCI-CCAP\_Mam6"  
/sex="female, virgin"  
/tissue type="infiltrating ductal carcinoma"  
/dev stage="5 months"  
/lab host="DH10B"  
/note="Organ: mammary; Vector: pCMV-Sport6; Site 1: SalI;  
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigator  
providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 150 a 179 c 128 g 148 t 1 others  
ORIGIN

Query Match 11.7%; Score 593.2; DB 10; Length 606;  
Best Local Similarity 99.3%; Pred. No. 9.7e-131;  
Matches 595; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 548 CACTTCCCTGTCATGAAGACAGCAACGAGTGTATTTCCAGTGCAGCAGCATTTGGTA 607  
Db 8 CGTTTCCCTGTCATGAAGACAGCAACGAGTGTATTTCCAGTGCAGCAGCATTTGGTA 67  
QY 608 TAGAGAGGATTTGCTAGATGCACCACTACAGCTGTGCTCTGTGCAATGGGCTCCA 667  
Db 68 TAGAGAGGATTTGCTAGATGCACCACTACAGCTGTGCTCTGTGCAATGGGCTCCA 127  
QY 668 TTCCAGTTGTTGAATGTCTCCACCTTAGTATCTTCTCCCTCCAGAGAGAGTGGCTC 727  
Db 128 TTCCAGTTGTTGAATGTCTCCACCTTAGTATCTTCTCCCTCCAGAGAGAGTGGCTC 187  
QY 728 CCTGGCTAAGGGGCTCTCCCTCATACAGCTGTGCTGTGGTGGCATCTGGCAG 787  
Db 188 CCTGGCTAAGGGGCTCTCCCTCATACAGCTGTGCTGTGGTGGCATCTGGCAG 247

QY 788 CCACATGGCCAAAGAGTTCCTTCAGCAGATCTGTGTCAGGCTCAATTAACACTA 847  
Db 248 CCACATGGCCAAAGAGTTCCTTCAGCAGATCTGTGTCAGGCTCAATTAACACTA 307  
QY 848 AGTTACAGAAAGGCATCATTCATTCGTCGCGCGCCCTCCAGAGCAGAAATGGGCTG 907  
Db 308 AGTTACAGAAAGGCATCATTCATTCGTCGCGCGCCCTCCAGAGCAGAAATGGGCTG 367  
QY 908 TGCTCAGTCATCAAGGCTGAATATACAGCAGTATTCACAGCAGCAACAAATATGA 967  
Db 368 TGCTCAGTCATCAAGGCTGAATATACAGCAGTATTCACAGCAGCAACAAATATGA 427  
QY 968 GTGACACTTACAGGTCGCCAGCAAGAACTTTCATGCAATCAAGTTTACCAGG 1027  
Db 428 GTGACACTTACAGGTCGCCAGCAAGAACTTTCATGCAATCAAGTTTACCAGG 487  
QY 1028 CTCAGCTGGCTTCAATATGGAATCTTCTGACATTTGATCAAGATGGAATCTGACG 1087  
Db 488 CTCAGCTGGCTTCAATATGGAATCTTCTGACATTTGATCAAGATGGAATCTGACG 547  
QY 1088 AAGAATTTATCTAGCTATGACCTAATTTGATGTGGCATGTGCTGACGCACTGCCG 1146  
Db 548 AAGAATTTATCTAGCTATGACCTAATTTGATGTGGCATGTGCTGACGCACTGCCG 606

Search completed: March 14, 2003, 11:05:19  
Job time : 6154.75 secs

GenCore version 5.1.4 p5 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2003, 01:40:05 ; Search time 8885.04 Seconds

(without alignments)  
11929.307 Million cell updates/sec

Title: US-09-674-237A-2

Perfect score: 3642  
Sequence: 1 atgctcagttccacacac.....tggaccaccagccagcaatga 3642

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

GenBml:\*

- 1: gb\_ba:\*
- 2: gb\_hcg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_ste:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vi:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_mu:\*
- 20: em\_om:\*
- 21: em\_or:\*
- 22: em\_ov:\*
- 23: em\_pat:\*
- 24: em\_ph:\*
- 25: em\_pl:\*
- 26: em\_ro:\*
- 27: em\_ste:\*
- 28: em\_un:\*
- 29: em\_vi:\*
- 30: em\_hcg\_hum:\*
- 31: em\_hcg\_inv:\*
- 32: em\_hcg\_other:\*
- 33: em\_hcg\_mus:\*
- 34: em\_hcg\_pln:\*
- 35: em\_hcg\_rod:\*
- 36: em\_hcg\_man:\*
- 37: em\_hcg\_vrt:\*
- 38: em\_sy:\*
- 39: em\_hcg\_hum:\*
- 40: em\_hcg\_mus:\*
- 41: em\_hcg\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3640.4	100.0	3723	10	AF132478
2	3639.4	99.9	5145	10	AF132481
3	3229.8	88.7	4025	10	AF127798
4	2831	77.7	5381	9	AF114488
5	2830	77.7	6439	9	AF114487
6	2827.8	77.6	5287	9	AF064243
7	2826.8	77.6	7247	9	AF064244
8	2816.2	77.3	3812	10	AF132672
9	2027.6	55.7	4103	5	AF032118
10	1259.2	34.6	2131	9	AK027846
11	986.4	27.1	3241	9	HS061166
12	749	20.6	1133	10	AF169621
13	721	19.8	5828	9	AF248540
14	721	19.8	5938	9	AB033082
15	713.8	19.6	4557	9	AF182199
16	713.8	19.6	6091	9	AF182198
17	677	18.6	4977	10	AF132479
18	673.8	18.5	3594	10	AF132480
19	648	17.8	4447	6	AX428899
20	531.4	14.6	78190	2	AC106760
21	531.4	14.6	179264	2	AC023112
22	477.4	13.1	2017	6	AX366305
23	387.4	10.6	1329	9	AK021545
24	342.2	9.4	1966	9	AF180522
25	332.4	9.1	2873	6	AF175271
26	331.4	9.1	1391	9	AF001630
27	331.4	9.1	2004	9	AF038189
28	327.8	9.0	171603	9	AC012629
29	258	7.1	747	6	AR175268
30	258	7.1	4053	9	HS061167
31	251	6.9	230097	2	AC126053
32	241.4	6.6	181782	2	AC116970
33	216.8	6.0	180637	2	AC098552
34	206.2	5.7	156481	2	AC123507
35	202.4	5.6	100000	9	AP000193
36	202.4	5.6	114929	9	AP000050
37	202.4	5.6	125242	9	AP000117
38	202.4	5.6	151516	9	AP000311
39	202.4	5.6	340000	9	AP001718
40	199.6	5.5	100000	9	AP000049
41	197.2	5.4	178801	2	AC121075
42	197.2	5.4	235426	2	AC121074
43	196	5.4	272545	2	AC090533
44	195.6	5.4	139113	2	AC129075
45	191.8	5.3	139113	2	AC129075

## ALIGNMENTS

RESULT 1  
AF132478  
LOCUS AF132478 3723 bp mRNA linear ROD 09-MAR-1999  
DEFINITION Mus musculus Esei protein mRNA, complete cds.  
ACCESSION AF132478  
VERSION AF132478.1 GI:4378884  
KEYWORDS  
SOURCE  
ORGANISM

Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE  
AUTHORS Sengar, A.S., Wang, W., Bishay, J., Cohen, S. and Egan, S.E.

TITLE The EH and SH3 domain Esei proteins regulate endocytosis by linking to dynamin and Eps15



```
QY 1501 CGACTGGCAACCCAGAGGCAAGAAATGAGAGCAGCAACAGTCTAGAGAGGCTAAGATTT 1560
DB 1501 CGACTGGCAACCCAGAGGCAAGAAATGAGAGCAGCAACAGTCTAGAGAGGCTAAGATTT 1560
QY 1561 GCTGAAATCAACCACTTACAGCAGCAGTTCAGAGAAATCTCAGCAAAATGCTTGGAGACTT 1620
DB 1561 GCTGAAATCAACCACTTACAGCAGCAGTTCAGAGAAATCTCAGCAAAATGCTTGGAGACTT 1620
QY 1621 ATTCCAGAGAAACAGATACCTACAGTACCACTTAAACAACTCCAGAGAAACAGTTCGAT 1680
DB 1621 ATTCCAGAGAAACAGATACCTACAGTACCACTTAAACAACTCCAGAGAAACAGTTCGAT 1680
QY 1681 AGAGACTCCGCTTCTTACCTCCAAAAGAGCCTTGAAGCAAGAGAGTGGCCCGCAGCAG 1740
DB 1681 AGAGACTCCGCTTCTTACCTCCAAAAGAGCCTTGAAGCAAGAGAGTGGCCCGCAGCAG 1740
QY 1741 CTCGGGAGCAGCTGAGCAGAGTGGAGAGAGACCAAGTCCAAAGCTGCAGAGATTTGAT 1800
DB 1741 CTCGGGAGCAGCTGAGCAGAGTGGAGAGAGACCAAGTCCAAAGCTGCAGAGATTTGAT 1800
QY 1801 GTTTTCAACCAACAGCTGAGAGAACTGAGAGATACATAGCAAAACAGCAATTCAGAG 1860
DB 1801 GTTTTCAACCAACAGCTGAGAGAACTGAGAGATACATAGCAAAACAGCAATTCAGAG 1860
QY 1861 CAGAGCTCCCTGAGGAGCGCCGACTGAAACAGAAAGAGAGAGAGAGAGAGCTTGGAG 1920
DB 1861 CAGAGCTCCCTGAGGAGCGCCGACTGAAACAGAAAGAGAGAGAGAGAGAGCTTGGAG 1920
QY 1921 TTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
DB 1921 TTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
QY 1981 GAGCATGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
DB 1981 GAGCATGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
QY 2041 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
DB 2041 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
QY 2101 GACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
DB 2101 GACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
QY 2161 CCCCTGCTTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
DB 2161 CCCCTGCTTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
QY 2221 GTATATTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
DB 2221 GTATATTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
QY 2281 CCAGAGATATATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAG 2340
DB 2281 CCAGAGATATATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAG 2340
QY 2341 GAGCTGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
DB 2341 GAGCTGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
QY 2401 GAGGTTCCCACTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2460
DB 2401 GAGGTTCCCACTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2460
QY 2461 GCTCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2520
DB 2461 GCTCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2520
QY 2521 AAGCACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2580
DB 2521 AAGCACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2580
QY 2581 GACAACTGGGATACGTGGGCGGCTCAGCTTCTCTGACCCGACTACTATGCTGGCAGTGA 2640
DB 2581 GACAACTGGGATACGTGGGCGGCTCAGCTTCTCTGACCCGACTACTATGCTGGCAGTGA 2640
QY 2641 CGGAGAGATCAGGCTTTTACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2700
DB 2641 CGGAGAGATCAGGCTTTTACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2700
QY 2701 GGCAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2760
DB 2701 GGCAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2760
QY 2761 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2820
DB 2761 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2820
QY 2821 ATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2880
DB 2821 ATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2880
QY 2881 CTATTTTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2940
DB 2881 CTATTTTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2940
QY 2941 GCTAGTCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3000
DB 2941 GCTAGTCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3000
QY 3001 ATTGCATGTACATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3060
DB 3001 ATTGCATGTACATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3060
QY 3061 GTGATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3120
DB 3061 GTGATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3120
QY 3121 GGAATCTTCCCTTCTTACATATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3180
DB 3121 GGAATCTTCCCTTCTTACATATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3180
QY 3181 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3240
DB 3181 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3240
QY 3241 ACTGTCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3300
DB 3241 ACTGTCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3300
QY 3301 CAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3360
DB 3301 CAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3360
QY 3361 TTTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3420
DB 3361 TTTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3420
QY 3421 CTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3480
DB 3421 CTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3480
QY 3481 GGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3540
DB 3481 GGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3540
QY 3541 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3600
DB 3541 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3600
QY 3601 TATGTAAAGCTGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660
DB 3601 TATGTAAAGCTGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3660
```



RESULT 2  
 AF132481 5145 bp mRNA linear ROD 09-MAR-1999  
 LOCUS Mus musculus Esell protein mRNA, complete cds.  
 DEFINITION AF132481  
 ACCESSION AF132481.1 GI:4378890  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mus musculus.  
 Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 10064583  
 2 (bases 1 to 5145)  
 Sengar, A.S., Wang, W., Cohen, S., Bishay, J. and Egan, S.E.  
 Direct Submission  
 Submitted (02-MAR-1999) Programs in Cancer & Blood  
 Research/Developmental Biology, The Hospital for Sick Children, 555  
 University Avenue, Toronto, ON M5G 1X8, Canada  
 Location/Qualifiers  
 1. 5145  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 1. 5145  
 /codon\_start=1  
 /product="Esell protein"  
 /protein\_id="AA19749.1"  
 /db\_xref="GI:4378891"  
 /translation="MAQPTPPGSGIDVWAIIVVERAKHDQFLSKPIAGTTGDQA  
 RMFFQSGLPQVLAQIMALADNNDQVEFSIAKLILKIQGQLPSTPPVM  
 KOQPAVISAAPAFIGISAMPPLTAVAPMGSIPIVGMSPPLSVPAVAPLAN  
 GAPVIOPLPAPAPATWPKSSFSRSGPSQNLTKOKQSPVPAAGAPAEVAP  
 OSSRLKROLFNSHDKTMSGHLGPQARTILMOSSLPOALQSLNLDIDQDLTA  
 EEFILAMHLIDVMSGQPLPVLPPEYIPSPRRRSGSGMSVSSSVDDRLPES  
 SEDBOPEKKLVVTEDEKRENFERSGVELEKRALLEQCKEERLADLEROEER  
 KEREQEOEAKROLEKQERLEKRELEKRELEKRELEKRELEKRELEKRELEK  
 NEROELNORKEOEGTVLAKARKTLEFELEKALDKHOLEKQDRECLARQOE  
 IESTNSRELRLAETIHLQOQLESQOMGLIPKQILSDQLKQONSJHPRSILT  
 LKRALEKELARQOLEOLDEVEREETRSKLOEDIVNNOLKREIRHSKQLOKORSI  
 EEARLKQKEERKSELEKQKDAORRQERDKMLHEVQOEORPKEPHEBRKX  
 EDSVKKEERAKPEMODKSRLEPHQERKLTQAPMSTTEGPTLISAQSVK  
 VYRRALYPPRESHDEITIQEDIVWDSQVTEGEMKQKGTGMPFPAVAKLP  
 ENEVTPAKPVTDLTISAPAKLRETPALRETPALRETPALRETPALRETPALRE  
 KETPNMDTWAQPSLTVPSAQLRQSRAPLPATATGSSPSPIVQKESVGLQAOL  
 YPMRAKDNHNLNFKNSDVTIVLEQODMMFGEVQCKMPKSYVKLISGVKRSKI  
 DTGPESPAISLRVASPAKPAIPGEELFAMTYTSSSGQDLPFOGGLIVVTKKSD  
 WMTGVGSGVGFPSNRYLNDSESGTAGKSGIKGKPELAIQAVTASAAAGPDLTI  
 AGGOLILIRKKNPGMVEGELQARCKQKQIGFPAHYKYLSPGSKITTPLEUKTAY  
 QPAVCOVIMDYTYKONDELAFSGQIINLVNKDPPMKGEVSGQGLPSPVVKL  
 TDMDPSSQWMSDHLDMLEPTEKRRQGYHELIVTEENVNDQLVTELFQKPLE  
 SLLLEKEVAMLFVWKEKLMCNILKALARKMSSEKMPVNMIGLISAQPHMQ  
 PYIRFSCQGLNGALIQOKTEADPKEFEVKRLAMPCKGKPLSSFLKMGVATRY  
 PLIKNILENTENHPDISHLEKAELEKELGVNREKENDREMTQAVHOC  
 GLSEOLVSNVNCIGPRKELSGKLYAKASKELVGLFEDFLILNOITPDLSSGT  
 DKVPSPKSLQYKMYKTPFLNEVYVKLPDPSGEPFLSHIDRVYTLAASINER  
 TAVOKIRASLELYLETEKKEKAYLVRSQKATIGLIMNVVGEELKRCRHSKS  
 NPEYCEVMSQCHITIKTIQDITLNPKNNSQCFIHLDEQVLCITVFERDQFSPDPL  
 GTEIRIVADIKKQDQSGKPVTKCLLHVEPTGEIVVRLDLQLEFEP"  
 BASE COUNT 1448 a 1298 c 1415 g 984 t  
 ORIGIN  
 Query Match 99.9%; Score 3639.4; DB 10; Length 5145;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 3640; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 61 GAAAGGCGCAAGCATGACACAGAGTTCCTTAGCCTGAAGCCGATAGGGGATTTTACT 120  
 DB 61 GAAAGGCGCAAGCATGACACAGAGTTCCTTAGCCTGAAGCCGATAGGGGATTTTACT 120  
 QY 121 GGTGATCAAGCGAGAACTTTTTCCTTCAATCTGGGTTACCTGAGCCTGCTTAGACCAA 180  
 DB 121 GGTGATCAAGCGAGAACTTTTTCCTTCAATCTGGGTTACCTGAGCCTGCTTAGACCAA 180  
 QY 181 ATATGGGCGCTAGGGAGACATGAATAGAGTAGAGTAGATCAAGTGAATTTTCCATA 240  
 DB 181 ATATGGGCGCTAGGGAGACATGAATAGAGTAGAGTAGATCAAGTGAATTTTCCATA 240  
 QY 241 GCGATGAAGCTTATCAAACTGAAGCTCAAGAGATATGAGTCCCTCCACACTTCCCTCCT 300  
 DB 241 GCGATGAAGCTTATCAAACTGAAGCTCAAGAGATATGAGTCCCTCCACACTTCCCTCCT 300  
 QY 301 GTGATGAAGACGCAACCAAGTGGCTATTTCAGTGCACACAGATTTTGGTATAGAGGATT 360  
 DB 301 GTGATGAAGACGCAACCAAGTGGCTATTTCAGTGCACACAGATTTTGGTATAGAGGATT 360  
 QY 361 GCTAGCATGACCACTACAGCTGTTGCTCTGCTGCAATGGGCTCCATTCCAGTTGTT 420  
 DB 361 GCTAGCATGACCACTACAGCTGTTGCTCTGCTGCAATGGGCTCCATTCCAGTTGTT 420  
 QY 421 GGAATGCTCCACCTTAGTATCTTCTGTCCTCCAGACAGAGTCCCTGCTGCTAAC 480  
 DB 421 GGAATGCTCCACCTTAGTATCTTCTGTCCTCCAGACAGAGTCCCTGCTGCTAAC 480  
 QY 481 GGGGCTCTCCGCTCATACAGCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
 DB 481 GGGGCTCTCCGCTCATACAGCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
 QY 541 AAGATGCTCTCTTACAGAGATCTGATCAAGGTCACAAATTAACACTTAATGACAGAG 600  
 DB 541 AAGATGCTCTCTTACAGAGATCTGATCAAGGTCACAAATTAACACTTAATGACAGAG 600  
 QY 601 GCAAAATATTCATGATGTCGCGCAGGCGCCCTCCAGACAGACAAATGGGCTGCTGCTGCTGCT 660  
 DB 601 GCAAAATATTCATGATGTCGCGCAGGCGCCCTCCAGACAGACAAATGGGCTGCTGCTGCTGCT 660  
 QY 661 TCAAGGCTGAATATCAGAGAGTATTAACAGACAGACAAATGAGTGAACACTTA 720  
 DB 661 TCAAGGCTGAATATCAGAGAGTATTAACAGACAGACAAATGAGTGAACACTTA 720  
 QY 721 ACAAGTCCCCAGGCAAGAACTATTCTGATCAATCAAGTTTACCAGGCTCACTGCTGCT 780  
 DB 721 ACAAGTCCCCAGGCAAGAACTATTCTGATCAATCAAGTTTACCAGGCTCACTGCTGCTGCT 780  
 QY 781 TCAATATGGAATCTTTCTGACATTTGATCAAGATGGAATACTCACTGCAAGAAATTTATC 840  
 DB 781 TCAATATGGAATCTTTCTGACATTTGATCAAGATGGAATACTCACTGCAAGAAATTTATC 840  
 QY 841 CTAGCTATGACCTAATATGATGTTGCCATGCTGCTGACGACCTGCGCCGCTGCTGCT 900  
 DB 841 CTAGCTATGACCTAATATGATGTTGCCATGCTGCTGACGACCTGCGCCGCTGCTGCTGCT 900  
 QY 901 CCAAAATACATCCCTGCTCTTCAAGAGAGTTCCTGCGGCGAGTGGATGTCCTGCTGCT 960  
 DB 901 CCAAAATACATCCCTGCTCTTCAAGAGAGTTCCTGCGGCGAGTGGATGTCCTGCTGCTGCT 960  
 QY 961 AGCTCTTCTCTGCTGATCAAGAGCTGCTGAGAGACCTGCTCAAGAGTGAAGTGAAGCAG 1020  
 DB 961 AGCTCTTCTCTGCTGATCAAGAGCTGCTGAGAGACCTGCTCAAGAGTGAAGTGAAGCAG 1020  
 QY 1021 CCAAGAGAAAGAACTGCTGCTGATGATTTGAAATGAAGAGGAGAACTTTGAGCAAGGC 1080  
 DB 1021 CCAAGAGAAAGAACTGCTGCTGATGATTTGAAATGAAGAGGAGAACTTTGAGCAAGGC 1080  
 QY 1081 AGTGTGAGCTGAGAGAAAGCGCCGCAAGGCTCTTGGAGCAGCAGGCGCAAGAGCAGAG 1140  
 DB 1081 AGTGTGAGCTGAGAGAAAGCGCCGCAAGGCTCTTGGAGCAGCAGGCGCAAGAGCAGAG 1140

QY 1141 CGGTGGCTCAGCTGAGCGCCGAGCAGAGAGGAAAGCGGAGCGCCAGAGCG 1200  
Db 1141 CGGTGGCTCAGCTGAGCGCCGAGCAGAGAGGAAAGCGGAGCGCCAGAGCG 1200  
QY 1201 GAGGCCAAGCGGAGCTGAGAGCTGAGAGAGCAGCTGAGAGAGAGCTGAGAGCG 1260  
Db 1201 GAGGCCAAGCGGAGCTGAGAGAGCAGCTGAGAGAGAGAGCTGAGAGCG 1260  
QY 1261 CAGCGAGAGAGAGAGAGAGAGATCGAGAGCGGAGCGCCAGAGAGAGCTG 1320  
Db 1261 CAGCGAGAGAGAGAGAGAGAGATCGAGAGCGGAGCGCCAGAGAGAGCTG 1320  
QY 1321 GAAAGCGAG 1380  
Db 1321 GAAAGCGAG 1380  
QY 1381 AACAG 1440  
Db 1381 AACAG 1440  
QY 1441 TTGAGAGCTCTGATGACAAAAAGATCAGCTAGAGAGAGAGAGAGAGAGAG 1500  
Db 1441 TTGAGAGCTCTGATGACAAAAAGATCAGCTAGAGAGAGAGAGAGAGAGAG 1500  
QY 1501 CGAAGGAG 1560  
Db 1501 CGAAGGAG 1560  
QY 1561 GCTGAAGTACCCATCTTACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620  
Db 1561 GCTGAAGTACCCATCTTACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620  
QY 1621 ATTCCAG 1680  
Db 1621 ATTCCAG 1680  
QY 1681 AGAGAGCTGCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 1740  
Db 1681 AGAGAGCTGCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 1740  
QY 1741 CTCGCGGAG 1800  
Db 1741 CTCGCGGAG 1800  
QY 1801 GTTTTCAACAG 1860  
Db 1801 GTTTTCAACAG 1860  
QY 1861 CAGAGGCTCTGAG 1920  
Db 1861 CAGAGGCTCTGAG 1920  
QY 1921 TTGAG 1980  
Db 1921 TTGAG 1980  
QY 1981 GAGCATGTGAG 2040  
Db 1981 GAGCATGTGAG 2040  
QY 2041 AAGAGGAG 2100  
Db 2041 AAGAGGAG 2100  
QY 2101 GAG 2160  
Db 2101 GAG 2160  
QY 2161 CCTGTGCTCAG 2220  
Db 2161 CCTGTGCTCAG 2220  
QY 2221 GTATATTAACGAG 2280

Db 2221 GTATATTAACGAG 2280  
QY 2281 CCAGAGAGATATAGCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340  
Db 2281 CCAGAGAGATATAGCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340  
QY 2341 GAGCTGAG 2400  
Db 2341 GAGCTGAG 2400  
QY 2401 GAGGTTCCACTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2460  
Db 2401 GAGGTTCCACTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2460  
QY 2461 GCTGTGAG 2520  
Db 2461 GCTGTGAG 2520  
QY 2521 AACAACTGGGAG 2580  
Db 2521 AACAACTGGGAG 2580  
QY 2581 GACAACTGGGAG 2640  
Db 2581 GACAACTGGGAG 2640  
QY 2641 CGGAG 2700  
Db 2641 CGGAG 2700  
QY 2701 GGCAG 2760  
Db 2701 GGCAG 2760  
QY 2761 AAAAG 2820  
Db 2761 AAAAG 2820  
QY 2821 ATGAG 2880  
Db 2821 ATGAG 2880  
QY 2881 CTCATTTTCAAG 2940  
Db 2881 CTCATTTTCAAG 2940  
QY 2941 GCTAGTCTTAAAG 3000  
Db 2941 GCTAGTCTTAAAG 3000  
QY 3001 ATTGCCATGTACATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3060  
Db 3001 ATTGCCATGTACATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3060  
QY 3061 GTGATTTGTGTTTAAAG 3120  
Db 3061 GTGATTTGTGTTTAAAG 3120  
QY 3121 GGAAGTCTTCCCTTCTTAACTATGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3180  
Db 3121 GGAAGTCTTCCCTTCTTAACTATGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3180  
QY 3181 AAAAG 3240  
Db 3181 AAAAG 3240  
QY 3241 ACTGATCCGAG 3300  
Db 3241 ACTGATCCGAG 3300  
QY 3301 CAGAGTGTGTGTGAG 3360

Db 3301 CCAGGTGATGTGGGAGAGAACTGCAAGCTCGAGGAGAAAAAGCCAGATAGGGTGG 3360  
 QY 3361 TTTCAGCAATATATGCAAACTTCTAAGCCCGGAACAAGCAAAATCAACCCCACTGAG 3420  
 Db 3361 TTTCAGCAATATATGCAAACTTCTAAGCCCGGAACAAGCAAAATCAACCCCACTGAG 3420  
 QY 3421 CTACCCAGAGCCGAGTGCAGGACGAGAGTGTGCGAGTGTATCGGATGTACATTACAC 3480  
 Db 3421 CTACCCAGAGCCGAGTGCAGGACGAGAGTGTGCGAGTGTATCGGATGTACATTACAC 3480  
 QY 3481 GCCAGAACATGACGACATGACGCTTCAAGAACGACATGATCATACGCTTCAACAG 3540  
 Db 3481 GCCAGAACATGACGACATGACGCTTCAAGAACGACATGATCATACGCTTCAACAG 3540  
 QY 3541 GAGAGCCGAGTGTGGGAGAGAGAGTCAATGAGGAGTGGCTCTTCCATCAAT 3600  
 Db 3541 GAGAGCCGAGTGTGGGAGAGAGAGTCAATGAGGAGTGGCTCTTCCATCAAT 3600  
 QY 3601 TATGTAAGTGAACACAGACATGAGCCCGAGCCAGCAATG 3641  
 Db 3601 TATGTAAGTGAACACAGACATGAGCCCGAGCCAGCAATG 3641

RESULT 3  
 AF127798 4025 bp mRNA linear ROD 28-JUN-1999  
 LOCUS Rattus norvegicus EH- and SH3-domain containing protein EHS1 mRNA,  
 DEFINITION complete cds.  
 ACCESSION AF127798  
 VERSION AF127798.1 GI:4835852  
 KEYWORDS  
 SOURCE Rattus norvegicus.  
 ORGANISM Rattus norvegicus.  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 4025)  
 Okamoto, M., Schoch, S. and Sudhof, T.C.  
 EHS1/intersectin, a protein that contains EH and SH3 domains and  
 binds to dynamin and SNAP-25. A protein connection between  
 exocytosis and endocytosis?  
 J. Biol. Chem. 274 (26), 18446-18454 (1999)

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 MEDLINE  
 PUBMED  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 FEATURES  
 source  
 CDS

1. 4025  
 /organism="Rattus norvegicus"  
 /db\_xref="taxon:10116"  
 41..3694  
 /codon\_start=1  
 /product="EH- and SH3-domain containing protein EHS1"  
 /protein\_id="AD30271.1"  
 /db\_xref="GI:4835852"  
 /translation="MAQPPFPGSLDIWAIVVERAKHDOFSLKPIISGITDQAR  
 RNFPSGSLPOPALQIALADLNKGRMDQVEFSIAKLITLKGQALPAPLPPVM  
 KQPPAIISSAPAFGIGMAGMPPPLTVAIVPMSIPVMSPLVSPQAAVPLAN  
 GAPPVQPLPAFAHPAATLPKSSFSRSPGSLNTKQKASFPVSAIPAAEWAIVP  
 OSRLRYQPLFNSHDKTMGHLTGQARITLQSSLPQALASINLSDIDGKLT  
 EEFILAMHLDIVMSGQPLPVLPEYIPSPRRVSGSGMSVSSSDQPLPEPS  
 SEDQOEKRLPTVEDKRENERGNLEKROALDQOEKOEERLQOEERLQOEER  
 KEREROERKRLQOEKOEERLQOEERLQOEERLQOEERLQOEERLQOEERLQOEER  
 NRQELLTORNDQEGIVLKARKTLEFELALNDKQKQLEGLQDIFCRALTOROE  
 ISTNKSRELRALIEITHLQOQESQMLGRLLPEKQILSDLKQVQNSLHRLDLT  
 LKRLAKELARQLEQLEDEKETRSLQEIIVFNNDLKLRLIHSKQQLQKRSI  
 EAERLKQOEERKSLLEKQEGRRVQOEERKQOEHEVQOEERKPRPHEDLTKR  
 EDSVKAEERAKPEVDQKSLFPHQEPKAPDAQMPPTTEKPLTISAQESKVV  
 YRALVPPFSRSHDEITIQPDIVMKGWDVESQIGRGWIGGPKKGTGMFANVA

EKIPENEIPPAKPVTLTSAAPKILAEETAPLPTVSSBSPSTPNMADSSWPS  
 STNEKETNDMDWAQPSLTVPSAQQLORSAFTPATRGSSPSPVLQGEKVBGLQ  
 AQLVYPRARKKDNHFNKSDVITVLEQDDHMLFEGVQDKMFPKSVKLSIGPVRK  
 STSIDGPTPEAPSSLEKRVASPAKPAIPEBEVAVMYTSSSHGLVTFQGVIVTK  
 KDGMWGTGVTGSGVFPNSNYLAKDSBSSTGKSLERKEBIAOVASTATGPE  
 QLTFLARQOLILIRKKNPGMWGELEOARSKQOIGMFPANVYKLSIPGSKITPTEL  
 KTAQVAPVQVIGMVDYTYQNDDELAFSGQILINVLKSDPMMWGEVSGQGLPSPN  
 YVKLITDMDPSQ"

BASE COUNT 1147 a 1009 c 1123 g 746 t  
 ORIGIN  
 Query Match 88.7%; Score 3229.8; DB 10; Length 4025;  
 Best Local Similarity 93.2%; Pred. No. 0;  
 Matches 3407; Conservative 0; Mismatches 232; Indels 18; Gaps 2;

QY 1 ATGGCTCATGTTCCCAACCTTTGGTGTGAGCTGATGTGGGCTATACCTGAGAG 60  
 Db 41 ATGGCTCATGTTCCCAACCTTTGGTGTGAGCTGATGTGGGCTATACCTGAGAG 100  
 QY 61 GAAGGGCCAGGACATGACGACAGCTTACGCTGAAGCCGATAGCGGGATTATTA 120  
 Db 101 GAAGAGCCAGGACATGACGACAGCTTACGAGTTTGAAGCCGATATGCAATCACT 160  
 QY 121 GGTGATCAAGGAGAACTTTTTCATCTGGATTACCTGACCTGTCTTAGACAA 180  
 Db 161 GGTGATCAAGGAGAACTTTTTCATCTGGATTACCTGACCTGTCTTAGACAA 220  
 QY 181 ATATGGGGCTAGGAGGACATGAATAAGATGGAAGATGATCAAGTGAATTTCCATA 240  
 Db 221 ATATGGGGCTAGGAGGACATGAATAAGATGGAAGATGATCAAGTGAATTTCCATA 280  
 QY 241 GCCATGAAGCTTATCAAACTGAAGCTAACGATATAGCTCCCTCCACACTTCCCT 300  
 Db 281 GCCATGAAGCTTATCAAACTGAAGCTAACGATATAGCTCCCTCCACACTTCCCT 340  
 QY 301 GTCATGAACAGCAACAGATGCTATTTCAGTGCACACAGCATTTGGTATAGAGAGATT 360  
 Db 341 GTCATGAACAGCAACAGATGCTATTTCAGTGCACACAGCATTTGGTATAGAGAGATT 400  
 QY 361 GCTGATGACGACAGCTGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420  
 Db 401 GCTGATGACGACAGCTGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 460  
 QY 421 GGAATGTCTCCACCTTATGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480  
 Db 461 GGAATGTCTCCACCTTATGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 520  
 QY 481 GGGGCTCTCCCTGATACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
 Db 521 GGGGCTCTCCCTGATACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 580  
 QY 541 AAGATTTCTTCTCAGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600  
 Db 581 AAGATTTCTTCTCAGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 640  
 QY 601 GACATATATTCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660  
 Db 641 GACATATATTCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 700  
 QY 661 TCAAGGCTGAATACAGGAGTATTCACAGCAGCAGCAAACTATAGTGAACATTA 720  
 Db 701 TCAAGGCTGAATACAGGAGTATTCACAGCAGCAGCAAACTATAGTGAACATTA 760  
 QY 721 ACAGGCTCCAGGAGCAAACTATTCATGATCAATCAAGTTTACCAGGCTCAGCTGGCT 780  
 Db 761 ACAGGCTCCAGGAGCAAACTATTCATGATCAATCAAGTTTACCAGGCTCAGCTGGCT 820  
 QY 781 TCAATATGGAATCTTTCTGATGATGATCAATGAGTGAAGAACTGACGAGAAATTTATC 840  
 Db 821 TCAATATGGAATCTTTCTGATGATGATGAGTGAAGAACTGACGAGAAATTTATC 880  
 QY 841 CTAGCTAAGCACTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900



QY 3046 TTTCAGCAAGGAGTGTGTTGTTACCAAGAAAGATGTGTCGTTGAGACGGGAGC 3105  
 DB 3098 TTTCAGCAAGGAGTGTGTTGTTACCAAGAAAGATGTGTCGTTGAGACGGGAGC 3157  
 QY 3106 GTGGGAGCAAGTCCGGAGTCTTCCCTTCTAATGTAGAGCTTAAAGATTGAGAGGC 3165  
 DB 3158 GTGGGAGCAAGTCCGGAGTCTTCCCTTCTAATGTAGAGCTTAAAGACTAGAGGC 3217  
 QY 3166 TCTGGAATCTGTGGGAAAAGAGGAGTTTGAAGAAAAAACCCTTAATTTGCCAGGTTATT 3225  
 DB 3218 TCTGGAATCTGTGGGAAAAGAGGAGTTTGAAGAAAAAACCCTTAATTTGCCAGGTTATT 3277  
 QY 3226 GCTTCTACCGCTGTACTGTCTCCGAAACAACCTGAGCTGCTGAGGAGCTGATTCTG 3285  
 DB 3278 GCTTCTACCGCTGTACTGTCTCCGAAACAACCTGAGCTGCTGAGGAGCTGATTCTG 3337  
 QY 3286 ATCCGAAAAAGAACCCAGGTGATGTGGGAGGAACTGCAAGCTGAGGAAAAAG 3345  
 DB 3338 ATCCGAAAAAGAACCCAGGTGATGTGGGAGGAACTGCAAGCTGAGGAAAAAG 3397  
 QY 3346 CGCCAGATAGGGTGTGTTTCCAGCAATTAATGCAAACTTCTAGCCCGGAAACAAGCAA 3405  
 DB 3398 CGCCAGATAGGGTGTGTTTCCAGCAATTAATGCAAACTTCTAGCCCGGAAACAAGCAA 3457  
 QY 3406 ATCACCCTCACTGAGCTACCCCAAGCCGAGTGCAGCACTGTGCGAGGTGATCGGG 3465  
 DB 3458 ATCACCCTCACTGAGCTACCCCAAGCCGAGTGCAGCACTGTGCGAGGTGATCGGG 3517  
 QY 3466 ATGTAGCATTAACACCCGCAAGCATGAGTAAGTACCTTTCAGCAAGGCGGAGTATC 3525  
 DB 3518 ATGTAGCATTAACACCCGCAAGCATGAGTAAGTACCTTTCAGCAAGGCGGAGTATC 3577  
 QY 3526 AACGCTCTCAACAGAGAGACCCGAGCTGTGGGAAAGGAAAGTCTGAGGAGTTGGG 3585  
 DB 3578 AACGCTCTCAACAGAGAGACCCGAGCTGTGGGAAAGGAAAGTCTGAGGAGTTGGG 3637  
 QY 3586 CTCTTCCATCAATTAATGTAAGCTGACCAAGACATGAGACCCGAGGAGCATTA 3642  
 DB 3638 CTCTTCCATCAATTAATGTAAGCTGACCAAGACATGAGACCCGAGGAGCATTA 3694  
 RESULT 4  
 AF114488 5381 bp mRNA linear PRI 16-JUL-2002  
 LOCUS AF114488 Homo sapiens intersectin short isoform (ITSN) mRNA, complete cds.  
 DEFINITION AF114488  
 ACCESSION AF114488.1 GI:4808824  
 VERSION AF114488.1  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 5381)  
 AUTHORS Arbones, M.L., Fuentes, J.J., Casas, C., de la Luna, S., Alcantara, S.,  
 Alu-splice cloning of human Intersectin (ITSN), a putative  
 multivalent binding protein expressed in proliferating and  
 differentiating neurons and overexpressed in Down syndrome  
 Eur. J. Hum. Genet. 7 (6), 704-712 (1999)  
 JOURNAL MEDLINE  
 MEDLINE 99415290  
 PUBMED 10482960  
 REFERENCE 2 (bases 1 to 5381)  
 AUTHORS Pucharcos, C., Fuentes, J.J., Pritchard, M. and Estivill, X.  
 Direct Submision  
 TITLE Submitted (16-DEC-1998) Medical and Molecular Genetics Center,  
 Cancer Research Institute, L'Hospitalet de l'Ho., Avia,  
 Castelldefels km. 2,7, Barcelona 08907, Spain  
 JOURNAL  
 FEATURES  
 Location/Qualifiers  
 1..5381  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="21"  
 /map="21q22.1-q22.2"

gene 1..5381  
 /gene="ITSN"  
 CDS 269..3931  
 /gene="ITSN"  
 /codon\_start=1  
 /product="intersectin short isoform"  
 /protein\_id="IAD29953.1"  
 /db\_xref="GI:4808825"  
 /translation="MAQPPPPGSGSLDIWATVEERAKHDQPHSLKPSISFTIGDOA  
 /RNFPOSGLPQVLAQIWAALNDNRNDQVDFEFLAKMLIKLQGLDLPALPVM  
 KOQPAVISAAPAFMGGLASMPPLTAVALAPVPGSLPVGMSPTLVSSVTAVALP  
 GAPPVIOPLPAFAHPAATLPKSSFSFGSGQNLTKLOKASPVAPVAPVAPV  
 OSSRLYKQPLFNSHDKMGHLTPQARTILMOSSLPQALSLINLSDIDDGLTA  
 BEFLIMHIDVMSGGQPLPVLPPEPIPPSPRRVSGSISVISTSDQLEPPEV  
 LEHFOOLKKLPPPEPKRRNPPENPENGILEKROALLLEQKROEAPOLERBOE  
 RKEREROBERKQLELEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLE  
 RNRQELNQRNREDEIVLAKKKTLEFELEALNDKQHLQLEGLQDRLTGLTORQ  
 EISTNKSEELRIAEITHLQOQLQESQMLGLIPEKQILNDQLEKQVQNSLHRSLSV  
 TLKRALAEKELARQHLRDQLEVEKETSRLQEIPIFNQLEKLEHNNKQLOKOS  
 MEARLKOKEOEKRIIELEKOEKOEKOEKOEKOEKOEKOEKOEKOEKOEKOEKOEK  
 KRESVKKKDKGEKOEKOEKOEKOEKOEKOEKOEKOEKOEKOEKOEKOEKOEKOEKOEK  
 KYVYRALYPFESRSHDETIOGDIVMKGMVDESOTGEPGMLGGELKGTGMPPA  
 NYAEKIPENEPAPVPKPTVDSAPAPKALAEPTPLAVTSSESTTPNNMADSSST  
 WPTSTNEKPEPTDWDMAAQPSLTVPSAQLEKQSAFPAIATGSSPSVLQGEVE  
 GLQQAQLYPMRAKNDHNLNKNNDVITVLEQDDMMWFGEVQKQMPKSVYKLISGP  
 IRKSTSDSGSSSPASLRVASPAKPVSESEFLAMTYSESSEGLTFOQGVIL  
 VTKSDDMWTGTGDKAGVPSNYVRLKDSGSGTAGTSGKGLKQVLAQVASYAT  
 GPQRLTALPQOLILIRKNPQGMWEGELQARKKQIGWFPANYVYKLISPGSTLTP  
 EPRKSTALAVCOVIGMYDTYTAONDELAFNKQIINVINKEDPDMWKEVNVQGLF  
 PSNYVLTITMDPSQO"  
 BASE COUNT 1629 a 1192 c 1304 g 1256 t  
 ORIGIN  
 Query Match 77.7%; Score 2831; DB 9; Length 5381;  
 Best Local Similarity 86.5%; Pred. No. 0;  
 Matches 3167; Conservative 0; Mismatches 475; Indels 21; Gaps 3;  
 QY 1 ATGCTCAGTTTCCACACCTTTGGTGTGAGCCTGATGTTGGGCATTACTGTGAG 60  
 DB 269 ATGCTCAGTTTCCACACCTTTGGTGTGAGCCTGATGTTGGGCATTACTGTGAG 328  
 QY 61 GAAGGGGCAAGCATGACGACGAGTCTTACGCTGAGCCGATGAGGGGATTTACT 120  
 DB 329 GAAGGGGCAAGCATGACGACGAGTCTTACGCTGAGCCGATGAGGGGATTTACT 388  
 QY 121 GGTATCAAGGAGGAATTTTTCCTCAATCTGGTGTACCTGAGCCTGTCTTACGACA 180  
 DB 389 GGTATCAAGGAGGAATTTTTCCTCAATCTGGTGTACCTGAGCCTGTCTTACGACA 448  
 QY 181 ATATGGGCGCTAGCGGACATGATTAAGATGAAGATGATGATCAAGTGAATTTTCCATA 240  
 DB 449 ATATGGGCGCTAGCGGACATGATTAAGATGAAGATGATGATCAAGTGAATTTTCCATA 508  
 QY 241 GGCATGAAGCTTTATCAAACTGAAGCTCAAGATATCAGCTCCCTCCACACTTCCCT 300  
 DB 509 GGCATGAAGCTTTATCAAACTGAAGCTCAAGATATCAGCTCCCTCCACACTTCCCT 568  
 QY 509 GGTATGAAGCTTTATCAAACTGAAGCTCAAGATATCAGCTCCCTCCACACTTCCCT 568  
 DB 689 GGTATGAAGCTTTATCAAACTGAAGCTCAAGATATCAGCTCCCTCCACACTTCCCT 748  
 QY 301 GTCATGAAGCAAGCAAGTGTCTATTTCAGTGCACACACATTTGTGTATGAGAGGATT 360  
 DB 569 GTCATGAAGCAAGCAAGTGTCTATTTCAGTGCACACACATTTGTGTATGAGAGGATT 628  
 QY 361 GGTAGCATGACACACACACACACGCTGCTCGTGTGCATATGGGCTCATTCAGTTGTT 420  
 DB 629 GGTAGCATGACACACACACACACGCTGCTCGTGTGCATATGGGCTCATTCAGTTGTT 688  
 QY 421 GGAATGTCTCCACCTTAGATCTTCTGCTCTCCACACACAGTGCCTCCCTGGCTAAC 480  
 DB 689 GGAATGTCTCCACCTTAGATCTTCTGCTCTCCACACACAGTGCCTCCCTGGCTAAC 748  
 QY 481 GGGGCTCTCCCGTCATACAGCTCTGCTGCTGTTGGGATCTCTGAGGCAATGAGCCA 540  
 DB 749 GGGGCTCTCCCGTCATACAGCTCTGCTGCTGTTGGGATCTCTGAGGCAATGAGCCA 808









Db 269 ATGGCTCAAGTTTCCAAACACCTTTTGGTGGCAGCTGGATATCTGGGCCATTAACCTAGAG 328  
 QY 61 GAAAGGGCCAGCATGACCGACAGATTCTTAGCCTGAAGCCGATAGCGGATTTATCT 120  
 Db 329 GAAAGGCGAAGCATGATGACGAGTTCCATAGTTTAAAGCAATATCTGGATTCTACT 388  
 QY 121 GGTGATCAAGCGAGAACTTTTTTTCATCTGGTTACTCTGAGCTGTCTTAGACAA 180  
 Db 389 GGTGATCAAGCGAAGAACTTTTTTTCATCTGGTTACTCTGAGCTGTCTTAGACAA 448  
 QY 181 ATATGGGGCTAGCGGACATGATTAAGATGAGAGATGATCAAGTGAATTTTTCATA 240  
 Db 449 ATATGGGGCTAGCGGACATGATTAAGATGAGAGATGATCAAGTGAATTTTTCATA 508  
 QY 241 GCCATGAAGCTTATCAAACTGAAGCTACAAAGATATCACTCCCTCCACACTTCCCT 300  
 Db 509 GCTATGAAGCTTATCAAACTGAAGCTACAAAGATATCACTCCCTCCACACTTCCCT 568  
 QY 301 GTGATGAAGCAAGCAAGCTAGTGTATTTCCAGTGCACCGACATTTGGTATGAGAGGATT 360  
 Db 569 GTGATGAAGCAAGCAAGCTAGTGTATTTCCAGTGCACCGACATTTGGTATGAGAGGATT 628  
 QY 361 GCTAGATGCAACCACTCAAGCTGTGCTCCTGTCGCAATGGGCTCAATCCAGTTGTT 420  
 Db 629 GCTAGATGCAACCACTCAAGCTGTGCTCCTGTCGCAATGGGCTCAATCCAGTTGTT 688  
 QY 421 GGAATGTCTCAACCTTAGTATCTTCTGCTCCAGCAGCAGTAGTCCCTCCGCTAAC 480  
 Db 689 GGAATGTCTCAACCTTAGTATCTTCTGCTCCAGCAGCAGTAGTCCCTCCGCTAAC 748  
 QY 481 GGGGCTCTCCGCTCATPACAGCTCTGCTGCGTTTGGCAGTCTGACCCACATGGCCA 540  
 Db 749 GGGGCTCTCCGCTCATPACAGCTCTGCTGCGTTTGGCAGTCTGACCCACATGGCCA 808  
 QY 541 AAGAGTCTTCTTCCAGCAGATGATGGTCCAGGGTCAATTAACCTAAGTTTACGAAG 600  
 Db 809 AAGAGTCTTCTTCCAGCAGATGATGGTCCAGGGTCAATTAACCTAAGTTTACGAAG 868  
 QY 601 GCACATCAATTCAGATGTCGCGACGCGCTCCAGCAGCAGAAATGGGCTGTGCTCAGTCA 660  
 Db 869 GCACATCAATTCAGATGTCGCGACGCGCTCCAGCAGCAGAAATGGGCTGTGCTCAGTCA 928  
 QY 661 TCAGGCTGAATATCAGGCAAGTATTTCACAGCAGCAGCAAACTATGAGTGAACATTTA 720  
 Db 929 TCAGGCTGAATATCAGGCAAGTATTTCACAGCAGCAGCAAACTATGAGTGAACATTTA 988  
 QY 721 ACGAGTCCCGCAGCAGAACTATTCATGCAATCAAGTTTACCCAGGCTCAGCTGGCT 780  
 Db 989 ACGAGTCCCGCAGCAGAACTATTCATGCAATCAAGTTTACCCAGGCTCAGCTGGCT 1048  
 QY 781 TCATATGGAATTTCTGACATTTGATCAAGATGAAATCTCACTGCAGAAATTTATC 840  
 Db 1049 TCATATGGAATTTCTGACATTTGATCAAGATGAAATCTCACTGCAGAAATTTATC 1108  
 QY 841 CTAGCTATGCACTAAATGATGTCGATGTCGATGTCGACGCACTGCGCGCTCTGCT 900  
 Db 1109 CTAGCTATGCACTAAATGATGTCGATGTCGATGTCGACGCACTGCGCGCTCTGCT 1168  
 QY 901 CGAGATPACATCCCTCTCTTCTCAAGAGTTCGCTCCGCGAGTGGAGTTCCTGATCA 960  
 Db 1169 CGAGATPACATCCCTCTCTTCTCAAGAGTTCGCTCCGCGAGTGGAGTTCCTGATCA 1228  
 QY 961 AGCTCTTCTCTGATGATGAGAGTTCGCTGCGAGAGCCGCTCAAGAGATGAGCAGAG 1020  
 Db 1229 AGCTCTTCTCTGATGATGAGAGTTCGCTGCGAGAGCCGCTCAAGAGATGAGCAGAG 1288  
 QY 1021 C---CAGAGAGAACTGCTGTGACATTTGAAGATPAGAGCGGAGAACTTTCAGCGA 1077  
 Db 1289 CAATTGAGAAAGAAATTAAGCTGTGACATTTGAAGATPAGAGCGGAGAACTTTCAGCGA 1348  
 QY 1078 GGAAGTGTGAGTGTGAGAGAGCGCCGCAAGCGCTCTTGGAGAGAGCGCAAGAGAG 1137  
 Db 1349 GGAAGTGTGAGTGTGAGAGAGCGCGCAAGCGCTCTTGGAGAGAGCGCAAGAGAG 1408  
 QY 1138 GAGCGTTTGGCTCAGCTGAGAGCGCGCTGAGCAGAGAGAGAAAGCGGAGCGCCAGAG 1197  
 Db 1409 GAGCGCTTGGCTCAGCTGAGAGCGCGCTGAGCAGAGAGAGAGAGCGGAGCGCCAGAG 1468  
 QY 1198 CAGAGGCGCAAGCGGCGCTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1257  
 Db 1469 CAGAGGCGCAAGCGGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1528  
 QY 1258 CGGAGCGAG 1317  
 Db 1529 CGGAGCGAG 1588  
 QY 1318 CTGAGAAAGCGAGCAGCACTTTGAATGGGAAACGAACTCGGAGACAGAACTTCTGATCAG 1377  
 Db 1589 CTGAGAAAGCGAGCAGCACTTTGAATGGGAAACGAACTCGGAGAGAGAACTTCTGATCAG 1648  
 QY 1378 AGGAAAG 1437  
 Db 1649 AGGAAAG 1708  
 QY 1438 GAGTTAGAGCTCTGAATGACAAAGAGCATCAGCTAGAGAGAGAGAGAGAGAGAGAG 1497  
 Db 1709 GAGTTAGAGCTCTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1768  
 QY 1498 TGTGACTGCGAACCCAGAGCGAAGAAATTTGAGAGCAGAACTTATGAGAGCTAAGA 1557  
 Db 1769 TGTGACTGCGAACCCAGAGCGAAGAAATTTGAGAGCAGAACTTATGAGAGCTAAGA 1828  
 QY 1558 ATTGCTGAATCAACCCAGCTTACAGCAGCAGTTCAGAGAACTCAGCAAACTTGGAGAG 1617  
 Db 1829 ATTGCTGAATCAACCCAGCTTACAGCAGCAGTTCAGAGAACTCAGCAAACTTGGAGAG 1888  
 QY 1618 CTATTCAG 1677  
 Db 1889 CTATTCAG 1948  
 QY 1678 CATAGAGCTGCTTTCTTACCTCAAAAGAGCTTGAAGCAAAAGAGCTGCGCCGAG 1737  
 Db 1949 CATAGAGCTGCTTTCTTACCTCAAAAGAGCTTGAAGCAAAAGAGCTGCGCCGAG 2008  
 QY 1738 CAGCTCCGAG 1797  
 Db 2009 CAGCTCCGAG 2068  
 QY 1798 GATGTTTCAACAACCAAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1857  
 Db 2069 GATGTTTCAACAACCAAGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2128  
 QY 1858 AAGCAGAGTCCCTGAG 1917  
 Db 2129 AAGCAGAGTCCCTGAG 2188  
 QY 1918 GAGTTAG 1977  
 Db 2189 GAGTTAG 2248  
 QY 1978 CTGAGAGATGTCAGCAGAGAG--GAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2034  
 Db 2249 CTGAGAGATGTCAGCAG 2308  
 QY 2035 AAGCTGAAG 2094  
 Db 2309 AAGCTGAAG 2368  
 QY 2095 ATGCAAGCAG 2154  
 Db 2369 ATGCAAGCAG 2428  
 QY 2155 CAGGACCTCTGCTTACCAAG 2214  
 Db 2429 CAGGACCTCTGCTTACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2488

QY 2215 AAGGTATATTATCCGAGCGGTGTAACCCCTTTGAAATCCAGAAATGACGATGATCAC 2274  
 Db 2489 AAGGTATATTATCCGAGCGGTGTAACCCCTTTGAAATCCAGAAATGATGAAATCAC 2548  
 QY 2275 ATCCAGCGAGAGATATATGATCAT-----GGTGAATGAAAGCCAGATGGA 2319  
 Db 2549 ATCCAGCGAGAGATATATGATCATGTTAAAGGGGAAATGGGTGATGAAAGCCAACTGGA 2608  
 QY 2320 GAGCCAGATGAGCTTGAGAGAGAGCTGAAAGGGAAGCGGATGATTCCTGCAAT 2379  
 Db 2609 GAACCCGAGCTTGAGAGAGATGAAAGGGAAGCGGATGATTCCTGCAAT 2668  
 QY 2380 GAGAGAAAGTTCCAGAAATGAGTTCCCATCCAGCCAAACCACTGACCTGACA 2439  
 Db 2669 GAGAGAAAGTTCCAGAAATGAGTTCCCATCCAGCCAAACCACTGACCTGACA 2728  
 QY 2440 TCGACCCCTGCCCCAACTGCTGCGTGAAGACCCCTGCTCTTGGCAGTGA 2499  
 Db 2729 TCGACCCCTGCCCCAACTGCTGCGTGAAGACCCCTGCTCTTGGCAGTGA 2788  
 QY 2500 TCTGAGCTTCCACCAACCCCAACCTGAGGAGACTTCACTTCACTGAGCCAGC 2559  
 Db 2789 TCAAGGCTTCCACCAACCCCAACCTGAGGAGACTTCACTTCACTGAGCCAGC 2848  
 QY 2560 TCAAGGAG 2619  
 Db 2849 AGCAATGAG 2908  
 QY 2620 GTACCTAGTCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2679  
 Db 2909 GTTCAAGTCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2968  
 QY 2680 TCTTCCCTGAG 2739  
 Db 2969 TCTTCCCTGAG 3028  
 QY 2740 CTGTATCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2799  
 Db 3029 CTGTATCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3088  
 QY 2800 ACCGTTCTGAG 2859  
 Db 3089 ACCGTTCTGAG 3148  
 QY 2860 TTCCCAAGCTTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2919  
 Db 3149 TTCCCAAGCTTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3208  
 QY 2920 ACTGAGCTTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2979  
 Db 3209 ACTGAGCTTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3268  
 QY 2980 GCGATTCGCGAG 3039  
 Db 3269 GCGATTCGCGAG 3328  
 QY 3040 TTACCTTTGAG 3099  
 Db 3329 TTACCTTTGAG 3388  
 QY 3100 GGAACGCTGAG 3159  
 Db 3389 GGAACGCTGAG 3448  
 QY 3160 GAGAGCTTCTGAG 3219  
 Db 3449 GAGAGCTTCTGAG 3508  
 QY 3220 GTTATGCTTCTGAG 3279  
 Db 3509 GTTATGCTTCTGAG 3568  
 QY 3280 ATTCTGATCCGAG 3339

Db 3569 ATTTGATCCCAAAAAAGAACCCAGGTGATGGTGGAGAGAGAGTGCAGAGACGTTGG 3628  
 QY 3340 AAAAGCGCCAGATAGAGGTGTTCCAGCAAAATTAATGTCAAATTTAGCCCCGAGACA 3399  
 Db 3629 AAAAGCGCCAGATAGAGGTGTTCCAGCAAAATTAATGTCAAATTTAGCCCCGAGAG 3688  
 QY 3400 AGCAAAATCAACCCCACTGAGCTTACCAAGACCGAGTGCAGCAGAGTGTCCAGGTG 3459  
 Db 3689 AGCAAAATCACTCCACAGAGCCAGCTTAAGTCAACAGCATTAAGCGGAGTGCAGGTG 3748  
 QY 3460 ATGGGATGTACGATTAACCGCCAGAGATAGAGAACTTACAGCAAAAGGCGAG 3519  
 Db 3749 ATGGGATGTACGATTAACCGCCAGAGATAGAGAACTTACAGCAAAAGGCGAG 3808  
 QY 3520 ATCATCAACGCTCTCAACCAAGAGAGCCGAGTGTGAGAGAGAGAGTCACTGGGCA 3579  
 Db 3809 ATCATCAACGCTCTCAACCAAGAGAGCCGAGTGTGAGAGAGAGAGTCACTGGGCA 3868  
 QY 3580 GTTGGGCTTTCCCATTCATTAATTAAGTGAACCAAGATGAGATGAGATGAGATGAG 3639  
 Db 3869 GTTGGGCTTTCCCATTCATTAATTAAGTGAACCAAGATGAGATGAGATGAGATGAG 3928  
 QY 3640 TG 3641  
 Db 3929 TG 3930

RESULT 6  
 AF064243  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 5287)  
 Guipponi, M., Scott, H.S., Chen, H., Schebesta, A., Rossier, C. and  
 Antonarakis, S.E.  
 Two isoforms of a human intersectin (ITSN) protein are produced by  
 brain-specific alternative splicing in a stop codon  
 Genomics 53 (3), 369-376 (1998)  
 98017974  
 9799604  
 2 (bases 1 to 5287)  
 Guipponi, M., Scott, H.S., Chen, H., Schebesta, A., Rossier, C. and  
 Antonarakis, S.E.  
 Direct Submission  
 Submitted (05-MAY-1998) Genetics and Microbiology, CMU, 1 rue  
 Michel-Servet, Geneva 4 CH-1211, Switzerland  
 Location/Qualifiers  
 1..5287  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="21"  
 /map="21q22.1-q22.2"  
 /tissue\_type="brain"  
 /dev\_stage="fetus"  
 107..3769  
 /codon\_start=1  
 /product="intersectin short form"  
 /protein\_id="AAC78610.1"  
 /db\_xref="GI:3859853"  
 /translation="NAQPTTPPGSLDIWATVEBRKHDQPFSLKLTGLKQGLPALPVM  
 RNFVDSGLPQVLAQIWLADNDGRMDQVBFSLMKLTGLKQGLPALPVM  
 KOQPAVIAAPPGMGILASMPPLRAVAAPGSLPVGMSPTLVASVPAVPAVPAV  
 GAPVIOPLPAFAPPAATLPKSSFSRSGSGLNTKLOKASPDVASVPAVPAVPAV  
 OSRLKYLQFNHSDKTMGHLTGPAFTILQSSLPQQLASIMWLSIDDDGLKTLA  
 EEFILMHLIDVMSGQPLPVLPEYIPSPRRVRSAGSIVISSTVDOLPEPVP  
 LDEQOOLEKKLPVTFEDKKRNFERNGLLEKRRQALLQGRKQERLAOLBRABOE



Db 1727 CTTATTCAGAAAAAGAGATCTCAATGACCAATTAAAACAAGTTCAGCAAGACGTTTG 1786  
 Qy 1678 CATAGAGACTCGCTTCTTACCTCAAAAGAGCCTTGAGACAAAGAGCGGCCGAG 1737  
 Db 1787 CAGAGATTCATCTTGTACACTTAAAGAGCCTTGAGAACAAAGAACTAGCTGGCAG 1846  
 Qy 1738 CAGCTCCGGAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1797  
 Db 1847 CACCTACGAGACCACTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1906  
 Qy 1798 GAGTTTTCAGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1857  
 Db 1907 GATATTTTCATATATACGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1966  
 Qy 1858 AAGCAGAGAGTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1917  
 Db 1967 AAGCAAAAGTCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2026  
 Qy 1918 GAGTTTTCAGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1977  
 Db 2027 GATATTTTCATATATACGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2086  
 Qy 1978 CAGGAGAGTGTGAT 2034  
 Db 2087 CAGGAGAGTGTGAT 2146  
 Qy 2035 AGACTGAT 2094  
 Db 2147 AAGCTGAT 2206  
 Qy 2095 ATGCAAGAT 2154  
 Db 2207 GAGCAAGAT 2266  
 Qy 2155 CAGGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2214  
 Db 2267 CAGGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2326  
 Qy 2215 AAGTGTGATATATACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2274  
 Db 2327 AAGTGTGATATATACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2386  
 Qy 2275 ATGCAAGAGAGAGATATGATCAT-----GATGATGAGAGAGAGAGAT 2319  
 Db 2387 ATGCAAGAGAGAGATATGATCATGATGATGATGATGATGATGATGATGATGAT 2446  
 Qy 2320 GAGCAGAGATGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2379  
 Db 2447 GAGCAGAGATGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2506  
 Qy 2380 GAGCAAAAGATTCAGAAAGATGAGTTCACCTCCAGCAGAGAGAGAGAGAT 2439  
 Db 2507 GAGCAAAAGATTCAGAAAGATGAGTTCACCTCCAGCAGAGAGAGAGAT 2566  
 Qy 2440 TCTGCCCCCTGCCCCCAAACTGCTGCTGAGAGAGAGAGAGAGAGAT 2499  
 Db 2567 TCTGCCCCCTGCCCCCAAACTGCTGCTGAGAGAGAGAGAGAGAGAT 2626  
 Qy 2500 TCTGAT 2559  
 Db 2627 TCTGAT 2686  
 Qy 2560 TCAAGAT 2619  
 Db 2687 ACGAATGAT 2746  
 Qy 2620 GTACCTAGAT 2679  
 Db 2747 GTTCCAGAT 2806  
 Qy 2680 TCTTCCCAATCTTCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAT 2739  
 Db 2807 TCTTCCCAATCTTCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAT 2866

Qy 2740 CTGATCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2799  
 Db 2867 CTATATCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2926  
 Qy 2800 ACCGTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2859  
 Db 2927 ACCGTTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2986  
 Qy 2860 TTCCCAAGTCTTACGAGAGAGAGAGAGAGAGAGAGAGAGAT 2919  
 Db 2987 TTCCCAAGTCTTACGAGAGAGAGAGAGAGAGAGAGAGAGAT 3046  
 Qy 2920 ACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 2979  
 Db 3047 TCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 3106  
 Qy 2980 GCGATCCCGAGAGAGAGAGAGAGAGAGAGAGAGAT 3039  
 Db 3107 GCGATCCCGAGAGAGAGAGAGAGAGAGAGAGAGAT 3166  
 Qy 3040 TTAACTTTGAGAGAGAGAGAGAGAGAGAGAGAGAT 3099  
 Db 3167 TTAACTTTGAGAGAGAGAGAGAGAGAGAGAGAGAT 3226  
 Qy 3100 GGAACGAGAGAGAGAGAGAGAGAGAGAGAGAT 3159  
 Db 3227 GGAACGAGAGAGAGAGAGAGAGAGAGAGAT 3286  
 Qy 3160 GAGGAGAGAGAGAGAGAGAGAGAGAGAT 3219  
 Db 3287 GAGGAGAGAGAGAGAGAGAGAGAGAGAT 3346  
 Qy 3220 GTTATGCTTCTTACGAGAGAGAGAGAGAGAGAGAGAT 3279  
 Db 3347 GTTATGCTTCTTACGAGAGAGAGAGAGAGAGAGAGAT 3406  
 Qy 3280 ATTGATCCGAGAGAGAGAGAGAGAGAGAGAGAT 3339  
 Db 3407 ATTGATCCGAGAGAGAGAGAGAGAGAGAGAT 3466  
 Qy 3340 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAT 3399  
 Db 3467 AAAAAAGAGAGAGAGAGAGAGAGAGAGAGAT 3526  
 Qy 3400 AGCAAAATCACTCCAGAGAGAGAGAGAGAGAGAGAT 3459  
 Db 3527 AGCAAAATCACTCCAGAGAGAGAGAGAGAGAGAGAT 3586  
 Qy 3460 ATCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3519  
 Db 3587 ATCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3646  
 Qy 3520 ATCATCAAGTCTTCAAGAGAGAGAGAGAGAGAGAGAT 3579  
 Db 3647 ATCATCAAGTCTTCAAGAGAGAGAGAGAGAGAGAGAT 3706  
 Qy 3580 GTTGGAGTCTTCCATCAATTAATTAAGTGAAGAGAGAGAT 3639  
 Db 3707 GTTGGAGTCTTCCATCAATTAATTAAGTGAAGAGAGAT 3766  
 Qy 3640 TGA 3642  
 Db 3767 TGA 3769

RESULT 7  
 AF064244 AF064244 7247 bp mRNA linear PRI 21-NOV-1998  
 LOCUS Homo sapiens Intersectin long form mRNA, complete cds.  
 DEFINITION  
 ACCESSION AF064244  
 VERSION AF064244.1 GI:3859854  
 KEYWORDS

SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS 1 (bases 1 to 7247)  
Guipponi M., Scott, H.S., Chen, H., Schebesta, A., Rossier, C. and Antonarakis, S.E.  
TITLE Two isoforms of a human interectin (ITSN) protein are produced by brain-specific alternative splicing in a stop codon  
JOURNAL Genomics 53 (3), 369-376 (1998)  
MEDLINE 99017974  
PUBMED 9799604  
REFERENCE 2 (bases 1 to 7247)  
AUTHORS Guipponi M., Scott, H.S., Chen, H., Schebesta, A., Rossier, C. and Antonarakis, S.E.  
TITLE Direct Submission  
JOURNAL Submitted (05-MAY-1998) Genetics and Microbiology, CMU, 1 rue Michel-Servet, Geneva 4 CH-1211, Switzerland  
FEATURES  
source  
1. 7247  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22.1-q22.2"  
/tissue\_type="brain"  
/dev\_stage="fetus"  
107. 5272  
/product="interectin long form"  
/protein\_id="AAC78611.1"  
/db\_xref="GI:385985"  
/translation="MAQEPPTPGGSLDIWATTEVERAKHQDFSLKPIESGITGDQA  
RNFPEQSLGPVLAQIALADPNNDMDQVEFISAKLLKLGQSLQSLAPVPM  
KQOPVAISAPPGFWMGIAISMPPLTAVAPVPMGPTLVSVPVAPVAPLAN  
GAPVIOPLPAFPAHPAATLTPKSSFSRSGSOLNTKOKRQSDVNSVPVAPVAP  
OSSRLYKQLEPNSHDKTMSGHITGPOARTIMOSTLPAQASIMSDIDQDKLTA  
BEFLAMHIDVAMSGPPLPVPPEYTPPSFRVRSVSGISVISTSVDRLEBPY  
LEDEQOULEKLPVTEEDKRENERFERGNLEEKRAQLEQKQERLAQLEADEP  
REREROERKQLEKLEKQLEKRELEKRELEKRELEKRELEKRELEKRELEKRE  
NRROELNORKEQEDIVLAKKKTLEFELALNDKQLEKQLEKQLEKQLEKQLE  
EESTNKRELR IAEITHLQOOLQESQOOLGRILPERKQILNDOLKOVONSLHDSIV  
TLKRLAEKELAROLRDOLDEVEKETSQLEIDIPNNQLEKELREINKQOLKOKS  
MEARLKQERKITELEKQERKQERKQERKQERKQERKQERKQERKQERKQERK  
KREESVKKRDEKQERKQERKQERKQERKQERKQERKQERKQERKQERKQERK  
KVYVRYALYPPESRSHDITTOPGDIVWKGWVDESQTGPMWAGKPLTKGTWPPA  
NYAEKIPENVEVPAVPVTDSTSAAPKTLARETPAPLAPVSSSESTPPNNAWFSST  
WPTNSEKETNDMDMAAOPSLTVPSAGOLROBSAFTPATATGSSPSPVGQEKVE  
GLQOALYPMRAKKNHNLNFKNDVITTLTEQDMWREBVGQKMPFKSYVYKLSGP  
IRKSTSMDSSESSEPSALIKRVASPAKPVVSGEFTAMYTESSSQGLTVOQDVL  
VTKSDWMTGTIVGDKAGVFPNSVRLKDSGSGTAGTKSLGKKPELAQVIASTAT  
GPEQLTLAAGOLILIRKKNPGMWGELOARQKQKQIPANVYKLLNPGTSKITPT  
EPFKSTALAAVCQVIGDYTAONDELAFNKGQIINVINKEDPMWKEVNGVGGVGF  
PSNVYKLTDDMPDSQWCSDLHLDMLPTPRKRGQYTHEILVRENVYNDLOLVTEI  
FOKPLMESELLEKEVAMI FVWKEKELINCNTKLKALVRRKMSGEKMPVMTGDIS  
AOLPMOPPIRFCSRLNGAALIQQKTEADPEKEFVRLMDPCKMPSSTILPK  
MORVTRPPIIKNILENTPENHSHKHALEKEBELCSQVNEBKEKESDLEWI  
QAHVCEGLSEOLVNSVTCGPKPKFLYAKKNGKLEGLFENDFILLQITK  
PLGSSGTPKVESPKSNLOYKWKPIPLFNEVYVLPDPSDEPIFHSIHLDDYVTLR  
AESINERTAVQKIKAASELYIETEKKREKAYIVRSRAGTIGELMNVVEGLEIKR  
CSHGKSNPFCEVTVGSOCHITKTIQDLPNPKNSNCFPIRDSQEVLCITVERDO  
FSPDPLGRTLRVADIKKQDSKGPVTKCLLHEVPPGELVLRDLQLFDEP"

misc\_feature 3569..3748  
/note="encodes SH3 domain"  
misc\_feature 3836..4390  
/note="encodes SH3 domain"  
misc\_feature 4649..4819  
/note="encodes SH3 domain"  
misc\_feature 4895..5143  
/note="encodes SH3 domain"  
BASE COUNT 2156 a 1652 c 1747 g 1692 t  
ORIGIN  
Query Match 77.6%; Score 2826.8; DB 9; Length 7247;  
Best Local Similarity 86.4%; Pred. No. 0;  
Matches 3164; Conservative 0; Mismatches 477; Indels 21; Gaps 3;  
QY 1 ATGGCTAGTTCACACACCTTTGGTGTAGCTGTGATCTGTGGCCATTAAGTGTGAG 60  
DB 107 ATGGCTAGTTCACACACCTTTGGTGTAGCTGTGATCTGTGGCCATTAAGTGTGAG 166  
QY 61 GAAAGGGCCAAAGCATGACGACGAGTTCCTTAGCCTGAAGCCGATAGCGGATTTACT 120  
DB 167 GAAAGGGCCAAAGCATGATGACGAGTTCCTTAGTAAAGCCAAATCTGGATTCATTACT 226  
QY 121 GGTATCAAGGAGAACTTTTTCCTCATCTGGGTACTTCACTGCTGCTTACGACAA 180  
DB 227 GGTATCAAGGAGAACTTTTTCCTCATCTGGGTACTTCACTGCTGCTTACGACAG 286  
QY 181 ATATGGGGCTAGCGGACATGAATTAAGATGAAGATGATCAAGTGAATTTTCATA 240  
DB 287 ATATGGGACATGCTGACATGAATTAAGATGAAGATGATCAAGTGAATTTTCATA 346  
QY 241 GCGATGAAGCTTATCAATCAATGAGCTTCAAGATATCAAGTCCCTCCACACTTCCCT 300  
DB 347 GCTATGAAGCTTATCAATCAATGAGCTTCAAGATATCAAGTCCCTCCACACTTCCCT 406  
QY 301 GTCATGAAGCAAGCAAGTGGCTATTTCCAGTGCACACAGATTTGGTATAGAGAGATT 360  
DB 407 GTCATGAAGCAAGCAAGTGGCTATTTCCAGTGCACACAGATTTGGTATAGAGAGATT 466  
QY 361 GCTATGAAGCAAGCAAGTGGCTATTTCCAGTGCACACAGATTTGGTATAGAGAGATT 420  
DB 467 GCGATGAAGCAAGCAAGTGGCTATTTCCAGTGCACACAGATTTGGTATAGAGAGATT 526  
QY 421 GAAATGCTCCACCTTATGATATCTTGTCTCTCCACAGCAAGTCCCTCCCTGCTAAC 480  
DB 527 GAAATGCTCCACCTTATGATATCTTGTCTCTCCACAGCAAGTCCCTCCCTGCTAAC 586  
QY 481 GGGGCTCTCCCTGCTATACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
DB 587 GGGGCTCTCCCTGCTATACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 646  
QY 541 AAGAGTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600  
DB 647 AAGAGTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 706  
QY 601 GCACATCATTCATGATGTCGCGAGCGCCCTCCACAGCAGACAGATGGGCTGCTGCTGCTGCT 660  
DB 707 GCAAGTCTATTTGATGTCGCGAGCGCCCTCCACAGCAGACAGATGGGCTGCTGCTGCTGCT 766  
QY 661 TCAAGGCTGAAATACAGCAGATTTTCAACAGCCAGCAGAAACTATGAGTGAACCTTA 720  
DB 767 TCAAGGCTGAAATACAGCAGATTTTCAACAGCCAGCAGAAACTATGAGTGAACCTTA 826  
QY 721 ACAGGTCCTCCAGCAGACATCTTCTCATGCAATCAAGTTTACCCGAGGCTGCTGCTGCT 780  
DB 827 ACAGGTCCTCCAGCAGACATCTTCTCATGCAATCAAGTTTACCCGAGGCTGCTGCTGCT 886  
QY 781 TCAATATGGAATCTTTCTGACATGATCAAGATGAGAACTCACTGAGAGAAATTTATC 840  
DB 887 TCAATATGGAATCTTTCTGACATGATCAAGATGAGAACTCACTGAGAGAAATTTATC 946  
QY 841 CTAGCTATGACCTTAATGATGTTGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900









QY 541 AAGATTTCTTCTTCAGCAGATCTGGTCCAGGCTCACAATTAACACTAAGTTACGAG 600  
 Db 581 AAGAGTTCTTCTTCAGCAGATCTGGTCCAGGCTCACAATTAACACTAAGTTACGAG 640  
 QY 601 GCACAAATCATTCGATGTCGCGCAGCGCCCTCCAGCAGCAGATGSGCTGTGCTCAGTCA 660  
 Db 641 GCACAAATCATTCGATGTCGCGCAGCGCCCTCCAGCAGCAGATGSGCTGTGCTCAGTCA 700  
 QY 661 TCAGAGCTGAATATACAGCAGCTTATTCACAGCCACGACAAATCTATGATGAGCACTTA 720  
 Db 701 TCAGAGCTGAATATACAGCAGCTTATTCACAGCCACGACAAATCTATGATGAGCACTTA 760  
 QY 721 ACAGATCCCCAGGCAAGAACTATTCTCATGCAATCAAGTTTACCCAGGCTCAGCTGGCT 780  
 Db 761 ACAGATCCCCAGGCAAGAACTATTCTCATGCAATCAAGTTTACCCAGGCTCAGCTGGCT 820  
 QY 781 TCATAATGGAATCTTTCTGACATTTGATCAAGATGAAAACTCATGACAGAAATTTATC 840  
 Db 821 TCATCTGGAAATCTTTCCGACATTTGATCAAGATGAAAACTCATGACAGAAATTTATC 880  
 QY 841 CTAGCTATGACCTTAATGATGTTGGCATGTCTGGTCAGCCATGCCCCCTGCTGCT 900  
 Db 881 CTAGCATGACCTTAATGATGTTGGCATGTCTGGTCAGCCATGCCCCCTGCTGCTGCT 940  
 QY 901 CCAGATATACATCCCTCCTTCTCAGAAAGATTGCTCCGCGCAGTGGAGATGTCCTGCTATA 960  
 Db 941 CCAGATATACATCCCTCCTTCTTAAAGAGTTGCTCTGGCATGAGTGTGCTGCTGCTATA 1000  
 QY 961 AGCTCTTCTTCTGTGATCAGAGGCTGCTGAGAGCCCTGCTCAGAGATGAGCAGCAG 1020  
 Db 1001 AGCTCTTCTGCTCAGAACCCAGCGGCTGCCGAGAGCCCTGCTCAGAGATGAGCAGCAG 1060  
 QY 1021 CCAGAGAAAGAACTGCTGTGATCATTTGAAAGATTAAGAACCGGAGAACTTTCAGACGAGGC 1080  
 Db 1061 GTGGAAGAAAGCTGCTGTGATCATTTGAAAGATTAAGAACCGGAGAACTTTCAGACGAGGC 1120  
 QY 1081 AGTGTGAGCTGAGAGAGCGCCGCAAGGCTCTTGGACAGCAGCGCAAGAGCAGAG 1140  
 Db 1121 AACCTGAGCTGAGAGAGCGCCGCAAGGCTCTTGGACAGCAGCGCAAGAGAGCAGAG 1180  
 QY 1141 CGGTTGGCTCAGCTGAGAGCGCCGCAAGAGAGAGAAAGCGGAGCGCCAGAGCAG 1200  
 Db 1181 CGGCTGGCTCAGCTGAGAGCGCCGCAAGAGAGAGAAAGCGGAGCGCCAGAGCAG 1240  
 QY 1201 GAGGCGAAGCGGAGCTGAGAGCTGAGAAAGCAGCTGAGAGAGCGGAGCGGAGCGG 1260  
 Db 1241 GAGCGAAGAGAGAGCTGAGAGCTGAGAAAGCAGCTGAGAGAGCGGAGCGGAGCGG 1300  
 QY 1261 CAGCGAAGAGAGAGAGAGAGATGAGAGGCGCGAGCGCCGCAAAACGGGAACTG 1320  
 Db 1301 CAGAGAGAGAGAGAGAGAGAGATGAGAGGCGCGAGCGCCGCAAAACGGGAGCTC 1360  
 QY 1321 GAAAGCAGAGCAACTTGAATGGGAACCGGAGACAGAAATCTCTGAATCAGAG 1380  
 Db 1361 GAAAGCAGAGCAACTTGAATGGGAACCGGAGACAGAAATCTCTGAATCAGAGAG 1420  
 QY 1381 AACCAAGAGCAG 1440  
 Db 1421 AACCAAGAGCAG 1480  
 QY 1441 TTGAGAGCTTGAATGACAAAAAGCATCAGCTAGAGAGAAAACTTCAGAGATACAGGTGT 1500  
 Db 1481 TTGAGAGCTTGAATGACAAAAAGCATCAGCTAGAGAGAAAACTTCAGAGATACAGGTGT 1540  
 QY 1501 CGAGCTGGCAACCCAGAGGAGAAATTTGAGAGACGAAACAGCTTGAAGAGAGTAAAGATT 1560  
 Db 1541 CGAGCTGGCAACCCAGAGGAGAAATTTGAGAGACGAAACAGCTTGAAGAGAGTAAAGATT 1600  
 QY 1561 GGTGAATACACCACTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATT 1620  
 Db 1601 GGTGAATACACCACTTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATT 1660  
 QY 1621 ATTCCAGAGAAACAGATCTCAGTGAACGATTAAAAAGAGTCCAGAGACAGATTGCTCAT 1680

Db 1661 ATTCCAGAGAAACAGATCTCAGTGAACAGTTTAAAAAGATCCAGAGAAACAGATTGCTCAT 1720  
 QY 1681 AGAGATCTGCTTTCTTACCTCAAAAAGAGCTTGGAGCAAAAGAGAGTGGCCCGGAGCAG 1740  
 Db 1721 AGAGATCTGCTTTCTTACCTCAAAAAGAGCTTGGAGCAAAAGAGAGTGGCCCGGAGCAG 1780  
 QY 1741 CTCCGGAGCAGCTGAGAGAGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800  
 Db 1781 CTTCGAGAGCAGCTGAGAGAGAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1840  
 QY 1801 GTTTTCAACCAACAGCTGAG 1860  
 Db 1841 GTTTTCAACCAACAGCTGAG 1900  
 QY 1861 CAGAGTCCCTGAG 1920  
 Db 1901 CAGAGTCCCTGAG 1960  
 QY 1921 TTGAG 1980  
 Db 1961 TTGAG 2020  
 QY 1981 GAGCATGTGAG 2040  
 Db 2021 GAGCATGTGAG 2080  
 QY 2041 AAGAGGAG 2100  
 Db 2081 AAGAGGAG 2140  
 QY 2101 GACAG 2160  
 Db 2141 GACAG 2197  
 QY 2161 CCTGTGCTTCAACAG 2220  
 Db 2198 CCTGTGCTTCAACAG 2257  
 QY 2221 GTATATTAACGAG 2280  
 Db 2258 GTATATTAACGAG 2317  
 QY 2281 CCAGAGATATATGATCAT-----GATGATGAAAGCCAGAGAGAGAGAGAGAGAGAG 2325  
 Db 2318 CCAGAGATATATGATCATGAG 2377  
 QY 2326 GATGAGCTTGGAG 2385  
 Db 2378 GATGAGCTTGGAG 2437  
 QY 2386 AAGATTCAG 2445  
 Db 2438 AAGATTCAG 2497  
 QY 2446 CTTGCCCCCAAACTGGCTCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2505  
 Db 2498 CTTGCCCCCAAACTGGCTCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2557  
 QY 2506 CCTTCACAAACCCCAACAACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2565  
 Db 2558 CCTTCACAAACCCCAACAACTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2617  
 QY 2566 GAGAGGCAAG 2625  
 Db 2618 GAGAGGCAAG 2677  
 QY 2626 AGTGTGCTGAG 2685  
 Db 2678 AGTGTGCTGAG 2737  
 QY 2686 CATCTCTCTGCTGAG 2745

```

Db 2738 CCTTCTCCCTGCTGGGCGAGGCTGAAAAAGTGGAGGCTCCAAACCAACCCCTGTAT 2797
Qy 2746 CCTTGAAGAGCCAAAAAGACAACTTAATTTTAACAAGAGAGTATCATCCGTT 2805
Db 2798 CCTTGAAGAGCCAAAAAGACAACTTAATTTTAACAAGAGAGTATCATCCGTT 2857
Qy 2806 CTGGAACACAGACACATGTGTGTGTGTGTGAGAGTTCAAGTCAAGAGGTTGTTCCCC 2865
Db 2858 CTGGAACACAGACACATGTGTGTGTGTGTGAGAGTTCAAGTCAAGAGGTTGTTCCCC 2917
Qy 2866 AAGCTTACGTGAAGTCAATTTTCAAGGCGCCGTTAAGAAATCCACAAGATCGATACGTC 2925
Db 2918 AAGCTTACGTGAAGTCAATTTTCAAGGCGCCGTTAAGAAATCCACAAGATCGATACGTC 2977
Qy 2926 CCTTACGAAGTCTGTCTGTCTGTAAAGAGAGTGTCTCCCGCGCCCAAGCCAGCCATT 2985
Db 2978 CCTACGGAAGTCTCTGTCTGTCTGTAAAGAGAGTGTCTCCCGCGCCCAAGCCAGCCATT 3037
Qy 2986 CCGGAGAGAGTGTATTTCCATGTACATACAGAGATTTGAGCAAGAGATTTAAAC 3045
Db 3038 CCGGAGAG----- 3046
Qy 3046 TTTACAGAAAGGAGTGTATGTGTGTACCAAGAAAGATGTGATGTGACGGGAGCG 3105
Db 3047 ----- 3046
Qy 3106 GTGGGCGACAGTCCGAGTCTTCCCTTCACTATGTAGAGCTTAAAGATTCAAGAGCG 3165
Db 3047 ----- 3046
Qy 3166 TCTGGAAGTGTGGAAAAAGGAGGATTTAGAAAAAACTGAAATTTGCCAGTTATT 3225
Db 3047 -----GAAATTCGCCAGTCAATT 3064
Qy 3226 GCTTCTACGCTGCTACTGTGTCCGAAACACTGACCTGTGCTGTGGAGAGTGTATTCTG 3285
Db 3065 GCTTCTACGCTGCTACTGTGTCCGAAACACTGACCTGTGCTGTGGAGAGTGTATTCTG 3124
Qy 3286 ATCCGAAAAAGAACCCAGTGTGTGTGGAAGAGAACTGAAAGTGTGAGGAAAAAG 3345
Db 3125 ATCCGAAAAAGAACCCAGTGTGTGTGGAAGAGAACTGAAAGTGTGAGGAAAAAG 3184
Qy 3346 CCGCAGATGAGGTGTGTTCAGCAAAATTATGCAAACTTTAGCCCGGAAACAAACAA 3405
Db 3185 CCGCAGATGAGGTGTGTTCAGCAAAATTATGCAAACTTTAGCCCGGAAACAAACAA 3244
Qy 3406 ATACCCCAACTAGTACTACCAAGACCGACGACGACGACGACGACGATGTGCGG 3465
Db 3245 ATACCCCAACTAGTACTACCAAGACCGACGACGACGACGACGACGATGTGCGG 3304
Qy 3466 ATGTACGATTACACCGCCGACGAAGATGACGAATGACCTTACGCAAGGCCAGATCATC 3525
Db 3305 ATGTACGATTACACCGCCGACGAAGATGACGAATGACCTTACGCAAGGCCAGATCATC 3364
Qy 3526 AAGCTCTTCAACAAGAGAGACCGGAGTGTGTGAAAGAGAAATGCTGGGCAAGTTGGG 3585
Db 3365 AAGCTCTTCAACAAGAGAGACCGGAGTGTGTGAAAGAGAAATGCTGGGCAAGTTGGG 3424
Qy 3586 CTCTTCCATCCATTTATGTAAGCTGACACAGACATGAGACCCAGCCAGCATGA 3642
Db 3425 CTCTTCCATCCATTTATGTAAGCTGACACAGACATGAGACCCAGCCAGCATGA 3481

```

```

RESULT 9
AF032118 4103 bp mRNA linear VRT 13-NOV-1998
LOCUS Xenopus laevis interseectin mRNA, complete cds.
DEFINITION
ACCESSION AF032118
VERSION AF032118.1 GI:2642624
KEYWORDS
SOURCE Xenopus laevis.
ORGANISM Xenopus laevis.
Enkalyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE
1 (bases 1 to 4103)
AUTHORS Yamabhai,M., Hoffman,N.G., Hardison,N.L., McPherson,P.S.,
Castagnoli,L., Cesarani,G. and Kay,B.K.
TITLE Intersectin, a novel adaptor protein with two Eps15 homology and
five Src homology 3 domains
JOURNAL J. Biol. Chem. 273 (47), 31401-31407 (1998)
MEDLINE 99030416
PUBMED 9813051
REFERENCE
2 (bases 1 to 4103)
AUTHORS Hoffman,N.G., Hardison,N.L., Yamabhai,M. and Kay,B.K.
TITLE Direct Submission
JOURNAL Submitted (30-OCT-1997) Pharmacology, University of Wisconsin, 1300
University Ave, Madison, WI 53706-1532, USA
FEATURES
source
location/Qualifiers
1..4103
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/cell_type="oocyte"
/function="involved in endocytosis"
/note="EH domain and SH3 domain containing protein;
similar to Intersectin binding proteins Ibp1, encoded by
GenBank Accession Number AF057285, and Ibp2, encoded by
GenBank Accession Number AF057286, and mouse
Rab/Rip, encoded by GenBank Accession Number AF057287"
/codon_start=1
/product="intersectin"
/protein_id="AAC73068.1"
/db_xref="GI:2642625"
/translation="MAQFGPPFGNLDIAITVEERAKHDQFHLKPTAGYITGDQ
RNFELSGLPQPLVLAQIMADNNDNRDQLEFSLAMKILKLGQYPLSLPDM
LKQPVMPAAAVAGRCMSGIVGIPPAVAAPVPMPSIPVMSPPVSSPPVPLPM
GAPAVIQSHFPAFASATLPEKSSFRGSVAGSQTNTLQAGSDVPAFPLVEMVAP
SSRLKRLQFNSQDRTMSGNLTGPQARTLMOSLPSQSLATTWLSIDIDQKLTAE
EFLIAMLIDVAMSGQPLPILPEYIIPSFRRVRSGLSISVSVQRLPEEPE
EERPNADKRLPVTFEKRENFERNGLIELEKQALIEQKQERLADLERAEORR
ERRODQERKRODLEKQERLEKQERLEKQERLEKQERLEKQERLEKQERLEKQ
RQELNQRREBDIVLAKAKKTLFELEALNDKKHLEGLDIDRLTLRHEI
ESTNKSRELRITLTHQQLQESQQLGKMTPEKSLIDQLQVQNSLHPSLTLE
KRALTEIGRIQRLDQLEVEKETRAKQLEIDVFNQQLKELRYNQKQFQDQPE
TEKIKQELERKTSLEDKLKEEDRKRLBDQKLMQORVQKEERYNFQDEKEREES
VQCEVEKKEPEIOEKPKPFPQPEPKLGGQIPMNNTTERAPLTINQGVKVVYRAL
YPPDARSHDEITTEPGDILIMVDSQTEGPEGLGELKGTGMPANVYAEEMSPSPS
TTTPAETTKPTTPTVHAPSPVAPAAATNTSTNENMADSSPTPTNTTKVESDMNDT
WAAQPSLTVPASQQRQSAFTPATVYGSPPSVLQGEKVBGLQALYPMRAKDN
HLNFKNDVITVLEQDDMMFMFGEVQKGFPSYVYLISGLPRKSTSIDTSSEPA
SLKRVSSPAFPAIIGEBEYISMTYESNEGDLTPOQGLIIVIKKDGMWGTVEK
TGVPSPSYVAPKXSEAAAGSGGTSGKKEILOVILASVATAPBQULTIAPGQILIR
KKNPGWMBEGLDARGGKROIGWPPANVYVLLSGNTKSTPTEPKPTSLPPTCOYIG
MYDIAONDELAFPSKQIVINLVKEDPDMKELNGHVLPSNVYKLITMDPBOQ
FRUGVXPAGGIIPATGRRPFLTFPPRGPSSLFPAFQAPLVSVMIRFCTTARFPCD
MNVYINII"
misc_feature
253..498
/note="encodes Eha domain"
misc_feature
650..1101
/note="encodes Ehb domain"
misc_feature
2386..2571
/note="encodes SH3A domain"
misc_feature
2881..3057
/note="encodes SH3B domain"
misc_feature
3148..3324
/note="encodes SH3C domain"
misc_feature
3364..3358
/note="encodes SH3D domain"
misc_feature
3607..3786
/note="encodes SH3E domain"
BASE COUNT 1290 a 920 c 1029 g 856 t 8 others
ORIGIN
Query Match 55.7%; Score 2027.6; DB 5; length 4103;
Best Local Similarity 73.6%; Pred. No. 0;

```

Matches 2685; Conservative 4; Mismatches 918; Indels 39; Gaps 7;

QY 1 ATGGCTCAGTTTCCACACCTTTGCGTGTAGCTGATGTCTGGCCATTAACCTGTGAG 60  
 Db 193 ATGGCTCAGTTTGGAACTCCGTTTGGGGGTAATTTGGACATCTGGCCATTAACCGGTGAG 252

QY 61 GAAAGGGCCAGCATACCAAGCAGTTCTTAGCTGAAGCCCATACCGGATTTTACT 120  
 Db 253 GAAAGGGCTTAACATACCAAGCAGTTCTTAGCTGAAGCCCATACCGGATTTTACT 312

QY 121 GGTATCAAGCAGAACTTTTTCATCTGGGTACTTACGCTCTCTTAGCACA 180  
 Db 313 GGTATCAAGCAGAACTTTTTCATCTGGGTACTTACGCTCTCTTAGCACA 372

QY 181 ATATGGGCGGTAGCGACATGAATTAACGATGAAGATGATTAAGTGAATTTTCATA 240  
 Db 373 ATATGGGCGGTAGCGACATGAATTAACGATGAAGATGATTAAGTGAATTTTCATA 432

QY 241 GCCATGAAGCTTATCAAACTGAAGCTACAAAGATATCAGTCCCTCCACACTTCCCTC 300  
 Db 433 GCTATGAATTAATCAAGCTAAACCTACAAAGTACCCCTGCTCTATTTCTCCCTTC 492

QY 301 GTCTGAAAACAGCAACCAAGTGTCTAT-----TTCCAGTGCACAGCATTTGTATAGA 354  
 Db 493 AATATGTGTAAGCAACCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 552

QY 355 GGGATTTGCTAGCATGCAACCACTCACAGCTGTTGCTCTGCTGCTGCTGCTGCTGCTG 414  
 Db 553 GGAATGTTGGGCACTTCTCTCCCTAGCATGTTAGCCCTGCTGCTGCTGCTGCTGCTG 612

QY 415 GTTGTGAATGTCTCACACCTTAGTATCTTGTCTCTCCACAGCAGTGTCTCTCTG 474  
 Db 613 GTATGAGAAATGTCTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 666

QY 475 GCTAAGGGGCTCTCTCCGCTACACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 534  
 Db 667 TCTAAGGGGCTCTCTCCGCTACAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 723

QY 535 TGCCCAAGAGTCTCTCTTACAGCAGATGTGCTCAGGCTCAAACTTAACCTAAGTTA 594  
 Db 724 TTGCCAAGAGTCTCTCTTACAGCAGATGTGCTCAGGCTCAAACTTAACCTAAGTTA 783

QY 595 CAGAGGCAATCATTTGATGTGCGCAGGCGCTCTCCACAGCAGATGCGCTGTGCT 654  
 Db 784 CAGAAAGCTCAGTCTCTGATGTGCGCAGGCGCTCTCCACAGCAGATGCGCTGTGCT 843

QY 655 CAGTCAACAGGCTGAATACAGCAGTTATTCACAGCAGCAGCAAACTTAAGTGA 714  
 Db 844 TCGTCTTCAAGATTTGAATACAGCAGTTATTCACAGCAGCAGCAAACTTAAGTGA 903

QY 715 CACTTAACAGGCTCCAGGCAAGAACTATTCTCATGCAATCAAGTTTACCCAGGCTCAG 774  
 Db 904 AATTTAACAGGCTCTCAGGCAAGAACTATTCTCATGCAATCAAGTTTACCCAGGCTCAG 963

QY 775 CTGGCTTCAATAGGAATCTTTCTGACATTTGATCAAGATGGAATACTCATCTCAAGAA 834  
 Db 964 CTAGCTACATATGGAATCTTTCTGACATTTGATCAAGATGGAATACTCACTCAAGAA 1023

QY 835 TTTATCTAGCTATGACCTAATTTGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 894  
 Db 1024 TTTATTTAGCTATGACCTAATTTGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1083

QY 895 CTGCTTCCAGAAATCACTCCCTCTCTCTTCAAGAGATTTGCTGCGCAGTGGGATGTC 954  
 Db 1084 CTGCTTCCAGAAATCACTCCCTCTCTTCAAGAGATTTGCTGCGCAGTGGGATGTC 1143

QY 955 GTCATAGGCTCTTCTGCTGATTCAGAGGCTGCTGAGAGCGGCTGCTGAGAGGATGAG 1014  
 Db 1144 ATTATAGGCTCTTCTGCTGATTCAGAGGCTGCTGAGAGCGGCTGCTGAGAGGATGAG 1203

QY 1015 CAGCAGCCAGAGAAAGCTGCTGTGACATTTGAAGATTAAGAGCGGAGAACTTTCAG 1074  
 Db 1204 CAAAACGCGAGCAAAAAGCTCCAGTGAATTTGAAGACAAAGAAAGGAGAACTTTGAG 1263

QY 1075 CGAGCAGTGTGAGCTTGAAGAAAGCCGCCCAAGCCCTCTTGGAGACGACGCCCAAGAG 1134  
 Db 1264 CGAGAAAGCTTAGAAATTAAGAAAAGAGAGACAGGCACTTTTAAGAGAGAGCCCAAGAG 1323

QY 1135 CAGAGAGGTTGGCTCAGCTGAGCGCGCCGAGCAGAGAGAGAAAGAGCGGAGCCCGAG 1194  
 Db 1324 CAGAGAGGTTGGCTCAGCTGAGCGCGCCGAGCAGAGAGAGAAAGAGAGAGCGGAG 1383

QY 1195 GAGCAGAGAGCCAAAGCGCAGCTGAGCTGAGAAAGCAGCTGAGAAAGCAGCGGAGCTG 1254  
 Db 1384 GATCAAGAGCGGAGAGCGGAGCAAGACCTGAGAAAGACAGCTGAGAAAGCAAGAGAACTG 1443

QY 1255 GAGCGGAGCGGAG 1314  
 Db 1444 GAGCGGAGCGGAG 1503

QY 1315 GAACCTGAAAAGCGCAGCACTTGAATGGAAACGAAACCGAGACAGAACTCTGAAT 1374  
 Db 1504 GAACCTGAAAAGCGCAGCACTTGAATGGAAACGAAACCGAGACAGAACTCTTAAAT 1563

QY 1375 CAGAGAAACAG 1434  
 Db 1564 CAAAGGAATAG 1623

QY 1435 TTTGATTAAGAGCTGGAATGAACAAAAGCATCAGCTGAAGAGAAACCTTCAAGATATC 1494  
 Db 1624 TTTGATTAAGAGCTGGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGA 1683

QY 1495 AGGTGCACTGGCAACCAAG 1554  
 Db 1684 CGGTGCGCTCACTACTCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1743

QY 1555 AGAATTCGTAATTAACCCCTTAACAGCAGTTCAGGAACTCAGCAAAATGCTTGA 1614  
 Db 1744 AGAATTCGTAATTAACCCCTTAACAGCAGTTCAGGAACTCAGCAAAATGCTTGA 1803

QY 1615 AGACTTAATCCAGAGAAACAGATTAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 1674  
 Db 1804 AATTAATTCCTGAGAAACAGTCCCTTATGATCAATTAAGCAGGTTCAACAAACAGT 1863

QY 1675 TTGATATGAAGTCTGCTTCTTCAACCAAGAGCTTGAAGCAAAAGAGTGGCGCCG 1734  
 Db 1864 TTGATATGAAGTCTGCTTCTTCAACCAAGAGCTTGAAGCAAAAGAGTGGAGCC 1923

QY 1735 CAGCAGCTCCGAGAGAGCTGAGCAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1794  
 Db 1924 CAGCAGCTCCGAGAGAGCTGAGTGAATGAATGAATGAATGAATGAATGAATGAATGA 1983

QY 1795 ATTGATGTTTCAACCAACAGCTGAAGAGAACTGAGAGAGATCATAGCAACAGCACTC 1854  
 Db 1984 ATGAGCTATTTTAATTAACCAATTAAGAGAACTGAGAGAACTGAGCAACAGCAGTTC 2043

QY 1855 CAGAGAGAGAGTCCCTGAG 1914  
 Db 2044 CAGAAACAGAGAGCTTTGAACTGAGAGAGATCAAGCAGAGAGAGAGAGAGAGAGAG 2103

QY 1915 CTGAGATTTAGAGAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1974  
 Db 2104 AGCAGCTGAGATTAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2163

QY 1975 TGGCTGAGAGATGTGACAG 2034  
 Db 2164 TGGCAGAGATGCGCTGAG 2217

QY 2035 AGACTGAAG 2094  
 Db 2218 AAG 2271

QY 2095 ATGCAAGACAG 2154  
 Db 2272 ATCAG 2331

QY 2155 CAGGACCCCTGGTCTTACACAGAGAAAGCCGCTTACCATTTCTGCACAGAGAGTGA 2214  
 Db 2332 CAGATTCCTGGATGAATACGAGAAAGCCCTCTAACGAT---CAACGAGGCGCATGTG 2388  
 QY 2215 AAGGTGTATTTATTCAGGACGCTGTATCCCTTTTGAATCCAGAAATCAGATGATCCAC 2274  
 Db 2389 AAGGTGTATTTATTCAGGACGCTGTATCCCTTTTGAATCCAGAAATCAGATGATCCACT 2448  
 QY 2275 ATCCAGACGAGAGATATATGATGTGATGAAAGCCAGCTGTAGAGAGCAGATGAGCTT 2334  
 Db 2449 ATTTGAGCTGTGAGACATTTATCATGTGTGATGAAAGCCAAATCGAGAGGCTGTGGCTG 2508  
 QY 2335 GAGAGAGAGCTGAAAGGAGAGAGCGGATGTTCCCTGCAACTATGACAGAAAGATTCCA 2394  
 Db 2509 GAGAGTGAATGAAAGGAGAGAGCGGCTGTTCTGCGCAATATGACAGAGGAAATGCC 2568  
 QY 2395 GAAATGAGATTTCCACTCCAGCCAAACCAATGACCGATCTGACATCTGCCCTGCCCC 2454  
 Db 2569 GAGAGTGAATTTCCAGAGCACAATTAACAGAGAGAGAACTACAGCCAAACCACTGTC 2628  
 QY 2455 AAAGCTGCTCTGGTGAGACCCCTGCTCTTCCAGTGAACCTCTTCTGAGCCCTCCACA 2514  
 Db 2629 CATGTGGAC-----CTTCCCTGTTGACCTGACCTTTTACAAATACATCTACA 2679  
 QY 2515 ACCCCCAACACTGGGACAGATTCACTTCCAGTGCAGGAGCCAGAGCTCAAGAGAGAGCA 2574  
 Db 2680 AATTCCAACTAGGCGACAGCTTCACTCAAGCTGAGCCAACTAACCACTGATTAAGTA 2739  
 QY 2575 GAAAGGAGCACTGGATATCTGAGGCGGCTCAGCTTCTCTGACCTACTAGTGTGCTGC 2634  
 Db 2740 GAGAGTGAATCTGGATATCATGGAGCTCAACATCTTAACTGTTCCAGTGTGCT 2799  
 QY 2635 CAGTTACGAGAGATCAGCTTTACCCAGACAGCACTGGCTCTCCCATCTCCC 2694  
 Db 2800 CAGCAACAGACAGGCTCTGCTTTTACCCAGCACTGTACTGATCATCGCATCTCT 2859  
 QY 2695 GTCTGTGGCCAGGAGTGAAGAGGCTCAAGAGCGCAAGCCCTGTATCTCTGAGAGA 2754  
 Db 2860 GTACTGTGTCAGGAGTGAAGAGGCTTCAAGAGCCCTGTATCTCTGAGAGA 2919  
 QY 2755 GCCAAAGAGCAACCACTTAATTTTAAACAAAGTCAAGCTCACTCCGTTCTGAGAGAG 2814  
 Db 2920 GCAAAAGAGCAACCACTTAATTTTAAACAAAGTCAAGCTCACTCCGTTCTGAGAGAG 2979  
 QY 2815 CAGACATGTGTGTTGAGAGAGTTCAAGGTCAGAGAGGTTGTTCCCAAGCTTAC 2874  
 Db 2980 CAGCATATGTGTGTTGAGAGAGTTCAAGGTCAGAGAGGTTGTTCCCAATCTCTAT 3039  
 QY 2875 GTGAACTCATTTTCAAGGCGCCGTAGAGAAATCCCAAGCATGATCTGACCTACTGAA 2934  
 Db 3040 GTAAAGCTTATTCGCTTCACTAAGAAATCTTAAAGCATGATCTTCAAGCTTCTGAGAA 3099  
 QY 2935 AGTCTGTCTAGTTAAAGAGAGTGGCTTCCCGGCGCCCAAGCACTTCCCGAGAA 2994  
 Db 3100 AGCCGAGCGGTTTAAAGAGAGTCTTCAACAGCATTTTAAACCTTCCATCCAAAGAGAA 3159  
 QY 2995 GAGTTATTTCCATGATACATACAGAGATTTGAGAGAGAGATTTAACTTCTGAGAA 3054  
 Db 3160 GAATACATATCTATACATATGAGAGACAGAGAGAGAGATTTGACCTTTGAGAA 3219  
 QY 3055 GGGAGATGTATGTGTGTTACAGAGAAAGATGTGACTGTGAGCGGAAAGCGTGGCGAG 3114  
 Db 3220 GAGATGTGTGTGTGTTAAAGAGAGATGAGAGCTGGTGTGAGCTGGGAGAGAG 3279  
 QY 3115 AAGTCCGAGATCTTCCCTTAACTATGTAGAGCTTAAAGATTCAGAGGCTGTGAGACT 3174  
 Db 3280 AAGATCTGGGTTTCTTCCCTCAACTAGTGAAGCCCAAGAGCTGGAGGCGCTGAGAG 3339  
 QY 3175 GCTGGGAGAAACAGGAGATTTAGAGAAAGAACTGAAATTTCCAGGTTATTTGCTTCTAC 3234  
 Db 3340 GAGAGAAAGACAGAGATCTTGGAGAGAAACAGAAATTTCCAGGTTATTTGATCTTAT 3399  
 QY 3235 GCTGTACTGTGTCCGAGAAACCTCACCTGGCTCTGGGAGCTGATTTGATCCGAGAA 3294

Db 3400 GCAGCAACTCTCTTGAAGCTPAACCTGGCTCTGGCCAACTTATCTCATCCAGAA 3459  
 QY 3295 AAGAACCCAGGTGATGTGTGAGAGAGAACTCAAGCTTGAAGGAGAAAGCCCGAGATA 3354  
 Db 3460 AAGATCTCAGAGAGATGTGTGAGAGAGAGACTGCAAGCATGTGGAGAAAGCGGAGATT 3519  
 QY 3355 GGGTGTCTTCCAGCAATATATGTCAAACTTTAAGCCCGGAAACAGCAAAATCACCCCA 3414  
 Db 3520 GGATGTTTCCAGCAAAATATATTTAAGCTTTTGAAGCTTGTGATTAATAAACACGCCCA 3579  
 QY 3415 ACTAGCTACCCCAAGACCGAGAGTCCAGCAGAGTGTGCCAGGTATCCGAGATGATGAT 3474  
 Db 3580 ACKGAACCCCAACCAAGCTATGCTCTTCAACAGCTTCAAGCTTATCGGATGATGAT 3639  
 QY 3475 TACACCGCCCAAGAGATGAGCAACTAGCTTGAAGCAAGAGCCAGATCATCAAGCTCTC 3534  
 Db 3640 TATATCCGAGAAATGACAGAGATGTGCTTCAATGTAAGGAGAGATTAAGTGTGCTG 3699  
 QY 3535 AACAGAGAGACCCGAGCTGTGTGAAAGAGAAATCAAGTGGGCAAGTTGGCTCTTCCA 3594  
 Db 3700 AACAAAGAGATCCCGAGCTGTGTGAAAGAGAACTGAATGGCATGTTGCTGTTCCC 3759  
 QY 3595 TCCAATATATGTAAGCTGACCAAGAGATGAGACCCCAAGCAAT 3640  
 Db 3760 TCCAACTAAGTGAACCTGACCAAGAGATGATCTTGAAGCAAT 3805

RESULT 10  
 AK027846  
 LOCUS  
 DEFINITION  
 Homo sapiens cDNA FLJ14940 fis, clone PLACE1010942, highly similar  
 to Homo sapiens Intersectin long isoform (ITSN) mRNA.  
 ACCESSION  
 AK027846  
 VERSION  
 AK027846.1 GI:14042823  
 KEYWORDS  
 oligo capping; fis (full insert sequence).  
 SOURCE  
 Homo sapiens placenta cDNA to mRNA, clone\_1lb:PLACE1  
 clone:PLACE1010942.  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE  
 AUTHORS  
 1 Isegai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,  
 Nishikawa,T., Nagai,K., Sugano,S., Ishibashi,T., Fujimori,K.,  
 Tanai,H., Kimata,M., Watanabe,M., Hirooka,S., Ishii,S., Kawai,Y.,  
 Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahara,K.,  
 Masuho,Y. and Kanehori,K.  
 Masuho,Y. and Kanehori,K.  
 NEDO human cDNA sequencing project

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 Isegai,T. and Otsuki,T.  
 Direct Submission  
 Submitted (10-MAY-2001) Takao Isegai, Helix Research Institute,  
 Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 282-0812, Japan  
 (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)  
 NEDO human cDNA sequencing project supported by Ministry of  
 Economy, Trade and Industry of Japan; cDNA full insert construction;  
 Research Association for Biotechnology; cDNA library construction;  
 5'- & 3'-end one pass sequencing and clone selection; Helix  
 Research Institute (supported by Japan Key Technology Center etc.)  
 and Department of Virology, Institute of Medical Science,  
 University of Tokyo.

FEATURES  
 source  
 1..2131  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="PLACE1010942"  
 /tissue\_type="placenta"  
 /clone\_id="PLACE1"  
 /note="cloning vector: pMB18SFL3"

BASE COUNT  
 ORIGIN  
 691 a 448 c 572 g 420 t

Query Match 34.6%; Score 1259.2; DB 9; Length 2131;  
 Best Local Similarity 81.2%; Pred. No. 26-274;  
 Matches 1539; Conservative 0; Mismatches 243; Indels 114; Gaps 2;

1 AAGGCTCAGTTTCCGACCTTTGGGAGTCTGGATGTCGGGCGCAATCTGGAG 60  
 347 ATGGCTCAGTTTCCGACCTTTGGGAGTCTGGATGTCGGGCGCAATCTGGAG 406  
 61 GAAAGGCGCAAGCATGACGAGTTCTTACCTGAGCGGATGAGGAGTTTACT 120  
 407 GAAAGGCGCAAGCATGACGAGTTCTTACCTGAGCGGATGAGGAGTTTACT 466  
 121 GGTGATCAAGCGGAGCACTTTTTCGATCTGGGTTTCTGAGGCTGTTAGACA 180  
 467 GGTGATCAAGCGGAGCACTTTTTCGATCTGGGTTTCTGAGGCTGTTAGACA 526  
 181 ATATGGGCGCTAGCGACATGATTAACGATGAGAGTGGATCAATGATTTTCTA 240  
 527 ATATGGGCGCTAGCGACATGATTAACGATGAGAGTGGATCAATGATTTTCTA 586  
 241 GCCATGAGCTTATCAAACTGAAAGTCAAGATATCAGCTCCCTCCACTTCCCT 300  
 587 GCTATGAACTTATCAAACTGAAAGTCAAGATATCAGCTCCCTCCACTTCCCT 646  
 301 GTCATGAACAGGACCGAGTGGCTATTTCCAGTGCACGACATTTGTATAGAGG 360  
 647 GTCATGAACAGGACCGAGTGGCTATTTCCAGTGCACGACATTTT----- 691  
 361 GCTAGAGTGCACGACCTCAAGCTGTGCTCTGTGCCAATGGGCTCAATCCAGTT 420  
 692 ----- 691  
 421 GGAATGCTCCACCTTATGATCTTCTGTCCCTCCAGGAGAGTGGCTCCCTGGTAC 480  
 692 ----- GAGCTGTGCTCCCTCCCTGGTAC 715  
 481 GGGGCTCTCCGCTATACAGCCTCTGCTGCTGCTTGGGATCTTGCAGCAATGG 540  
 716 GGGGCTCTCCGCTTATACAGCCTCTGCTGCTGCTTGGGATCTTGCAGCAATGG 775  
 541 AAGAGTTCTTCTTACAGCAGATCTGCTCAAGGCTCAATTAACCTTAAGTTACA 600  
 776 AAGAGTTCTTCTTATGATATCTGATCTGATCTCAAGGCTCAATTAACCTTA 835  
 601 GCAATTCATTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660  
 836 GCAATTCATTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 895  
 661 TCAAGGCTGAAATACAGGAGTTTATCAACGCTGACGACAAATCTATGATGAC 720  
 896 TCAAGGCTGAAATACAGGAGTTTATCAACGCTGACGACAAATCTATGATGAC 955  
 721 AAGAGTTCTTCTTACAGCAGATCTGCTCAAGGCTCAATTAACCTTAAGTTACA 780  
 956 AAGAGTTCTTCTTACAGCAGATCTGCTCAAGGCTCAATTAACCTTAAGTTACA 1015  
 781 TCAATATGATCTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 840  
 1016 TCAATATGATCTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 1075  
 841 CTAGCTATGACCTTAATGATGATGATGATGATGATGATGATGATGATGATG 900  
 1076 CTAGCTATGACCTTAATGATGATGATGATGATGATGATGATGATGATGATG 1135  
 901 CCAAGATACATCT 960  
 1136 CCAAGATACATCT 1195  
 961 AGCT 1020  
 1196 AGCT 1255  
 1021 C---CAGAGAGAAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1077

1256 CAATTAAGAAAGAAATTAATCTGTAACCTTTGAAGATTAAGAGGAGAACTTTGAACG 1315  
 1078 GGCAGTGTGAGCTGAGAGAACCGCGCAAGCGCTCTTGGAGCAGAGCGCAAGAGCAG 1137  
 1316 GGCAGTGTGAGCTGAGAGAACCGCGCAAGCGCTCTTGGAGCAGAGCGCAAGAGCAG 1375  
 1138 GAGCGTTTGGCTCAGCTGAGAGCGCGCGAGCAGAGAGAGAGAGAGAGAGAGAGAG 1197  
 1376 GAGCGTTTGGCTCAGCTGAGAGCGCGCGAGCAGAGAGAGAGAGAGAGAGAGAGAG 1435  
 1198 CAGAGAGCGAGCGAGCTGAGAGCGCGCGAGCAGAGAGAGAGAGAGAGAGAGAGAG 1257  
 1436 CAGAGAGCGAGCGAGCTGAGAGCGCGCGAGCAGAGAGAGAGAGAGAGAGAGAGAG 1495  
 1496 CAGAGAGCGAGCGAGCTGAGAGCGCGCGAGCAGAGAGAGAGAGAGAGAGAGAGAG 1555  
 1318 CTGAGAGAGCGAGCGAGCTTGAATGGGAACTGGAGAGAGAGAGAGAGAGAGAGAG 1377  
 1556 CTTGAGAGCGAGCGAGCTTGAATGGGAACTGGAGAGAGAGAGAGAGAGAGAGAGAG 1615  
 1378 AG 1437  
 1616 AG 1675  
 1438 GAGTTAGAGCTCTGATTAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1497  
 1676 GAGTTAGAGCTCTGATTAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1735  
 1498 TGTGAGCTGAG 1557  
 1736 TGTGAGCTGAG 1795  
 1558 ATTGCTGAATACCCCTTACAGCAGAGCTTGCAGAGATCTTCAAGATCTTGAAGA 1617  
 1796 ATTGCTGAATACCCCTTACAGCAGAGCTTGCAGAGATCTTCAAGATCTTGAAGA 1855  
 1618 CTTATTCAG 1677  
 1856 CTTATTCAG 1915  
 1678 CATAGAGCTGCTCTTCTTACCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1737  
 1916 CACAGAGCTGCTCTTCTTACCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1975  
 1738 CAGCTCGGAG 1797  
 1976 CAGCTCGGAG 2035  
 1798 GATGTTTCAAG 1857  
 2036 GATGTTTCAAG 2095  
 1858 AAGCAGAGGTCCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1893  
 2096 AAGCAGAGGTCCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2131

RESULT 11  
 HSU61166  
 LOCUS HSU61166 3241 bp mRNA linear PRI 23-JUL-1996  
 DEFINITION Human SH3 domain-containing protein SH3P17 mRNA, complete cds.  
 ACCESSION U61166  
 VERSION U61166.1 GI:1438932  
 KEYWORDS  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 3241)  
 AUTHORS Sparks, A.B., Hoffman, N.G., McConnell, S.J., Fowlkes, D.M. and





QY	3471	CGATTACCGCCGACAAACGATACCAATAGAGCTTCAGCAAAAGGCGACATCATCAACGT	3530
Db	1428	CGACTACCGCGCAGAAATAGACATAGTACGTGGCTTCAACAAAGGCGACATCATCAACGT	1487
QY	3531	CCTCAACAAAGAGAGACCCCGAGCTGTGTGAAAGAGAAAGTCAAGTGGCAAGTTGGGCTCTT	3590
Db	1488	CTCAACAAAGAGAGACCCCTGAGCTGTGTGAAAGAGAAAGTCAAGTCAAGTGGGCTCTT	1547
QY	3591	CCCATCCATTATTTAGTAAGCTGACCAAGACATGAGACCCAGCAGACATGTA	3642
Db	1548	CCCATCAATTATGTAGCTGACCAAGACATGAGACCCAGCAGACATGTA	1599
RESULT 12	AF169621	1133 bp	mRNA linear
LOCUS	AF169621		
DEFINITION	Mus musculus Esei protein mRNA, partial cds.		
ACCESSION	AF169621		
VERSION	AF169621.1	GI:5731280	
KEYWORDS			
SOURCE	Mus musculus.		
ORGANISM	Mus musculus.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 1133) Tsyba,L.O., Kvasha,S.M., Skripkina,I.Y., Anoprienko,O.V., Slavov,D., Tassone,F., Rynditch,A.V. and Gardiner,K.		
TITLE	Mouse homologues of human chromosome 21 genes		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 1133) Tsyba,L.O., Kvasha,S.M., Skripkina,I.Y., Anoprienko,O.V., Slavov,D., Tassone,F., Rynditch,A.V. and Gardiner,K.		
AUTHORS	Direct Submission		
TITLE	Submitted (15-JUL-1999) Department of Molecular Oncogenetics, Institute of Molecular Biology and Genetics of National Academy of Sciences of Ukraine, 150, Zabolotnogo str., Kiev 252677, Ukraine		
FEATURES	location/Qualifiers		
source	1..1133		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/dev_stage="15-day embryo"		
	<1..749		
	/note="similar to Homo sapiens intersectin"		
	/codon_start=3		
	/product="Esei protein"		
	/protein_id="AAD4848.1"		
	/db_xref="GI:5731281"		
	/translation="VRKSTSIDTGPTESPASIKRVASPAKPAIPGEFIAMTYESS EGQDITFOOGDVIIVTKKDDMWITVADKSGVPSNVVR.LKDSGSGTAAKTSIGL KEPIQVIAAATGAEPEQLTLAGPGLIIRKNPGMWEGELQARGKROIQWPPANV VKLSRPSKTIPTPELPRKTAOPARCOVIGIKIDYTAONDELAPSKGQIINVUNKEDP DMWKEVSGQGLFSPSNVYKLTITDPEQQ"		
	DMWKEVSGQGLFSPSNVYKLTITDPEQQ"		
BASE COUNT	316 a 276 c 285 g 256 t		
ORIGIN			
Query Match	20.6%; Score 749; DB 10;	Length 1133;	
Best Local Similarity	100.0%; Pred. No. 8e-159;		
Matches 749; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	2894	CCGTAAAGAAATCCACAGACATCGATCTAGCCCTTACTGAAAGTCTGCTAGTCTAAAGA	2953
Db	1	CCGTAAAGAAATCCACAGACATCGATCTAGCCCTTACTGAAAGTCTGCTAGTCTAAAGA	60
QY	2954	GAGTGGCTTCCCCGGCCGACCAAGCCGACATTCGCCGAGAAAGATTATGCGACATGTACA	3013
Db	61	GAGTGGCTTCCCCGGCCGACCAAGCCGACATTCGCCGAGAAAGATTATGCGACATGTACA	120
QY	3014	CATACGAGAGTTCTGAGCAAGAGAGATTTAACCTTTAGCAAGAGGAGATGTATGTGGTTA	3073
Db	121	CATACGAGAGTTCTGAGCAAGAGAGATTTAACCTTTAGCAAGAGGAGATGTATGTGGTTA	180
QY	3074	CCAAGAAAGATGTGACTGTGTGACGGGAAACGTTGGGCAACAAGTCGAGATCTTCCCTT	3133

Db	181	CCAAAGAAAGTGTACTGTGTGACGGGAAACGGTGGGACGAACAAGTCCGAGTCTTCCCTT	240
QY	3134	CTAACTATGTGAGGCTTAAATTCAGAGGGCTCTGGAACTGTGGGAAACAGGAACTT	3193
Db	241	CTAACTATGTGAGGCTTAAAGATTTCAGGGGCTCTGGAACTGTGGGAAACAGGAACTT	300
QY	3194	TAGAAAAAAACCTGAAATTGCCAGTGTATTGTCTCTAGCGTCTACTGTTCCCGAAC	3253
Db	301	TAGAAAAAAACCTGAAATTGCCAGGTTATTGTCTCTAGCGTCTACTGTTCCCGAAC	360
QY	3254	AACTACCTTGGCTCTTGGGAGCTGATTCTGATTCGGAAAAAGAACCCAGGTGATGCT	3313
Db	361	AACTACCTTGGCTCTTGGGAGCTGATTCTGATTCGGAAAAAGAACCCAGGTGATGCT	420
QY	3314	GGGAAGGAGAACTGCAAGCTGAGGGGAAAAAGCCGACATGATGGGGTGTTCACGAAATT	3373
Db	421	GGGAAGGAGAACTGCAAGCTGAGGGGAAAAAGCCGACATGATGGGGTGTTCACGAAATT	480
QY	3374	ATGTCAAACTTTAAAGCCCGGAAACAGCAAAATATCAACCCCAACTGAGTACCAAGACCG	3433
Db	481	ATGTCAAACTTTAAAGCCCGGAAACAGCAAAATATCAACCCCAACTGAGTACCAAGACCG	540
QY	3434	CAGTCACACCGACGAGTGTGCCAGGATCCGGAGTCAAGATTACACGCCCAAGACGATG	3493
Db	541	CAGTCACACCGACGAGTGTGCCAGGATCCGGAGTCAAGATTACACGCCCAAGACGATG	600
QY	3494	ACGAACCTGACCTTCAGCAAAAGCCAGATCATCAACGTCCTCAACAGAGAGACCCGAGCT	3553
Db	601	ACGAACCTGACCTTCAGCAAAAGCCAGATCATCAACGTCCTCAACAGAGAGACCCGAGCT	660
QY	3554	GGTGAAGAGGAGAACTCAGTGTGGGCAAGTTGGGCTTTCCCATTCATATGTAAAGCTGA	3613
Db	661	GGTGAAGAGGAGAACTCAGTGTGGGCAAGTTGGGCTTTCCCATTCATATGTAAAGCTGA	720
QY	3614	CCACAGACATGAGACCCGACGACGACATGA	3642
Db	721	CCACAGACATGAGACCCGACGACGACATGA	749

RESULT 13	AF248540	5828 bp	mRNA	linear	PRI 12-Apr-2000
LOCUS	AF248540				
DEFINITION	Homo sapiens intersecin 2 (SH3D1B) mRNA, complete cds.				
ACCESSION	AF248540				
VERSION	AF248540.1	GI:7542782			
KEYWORDS					
SOURCE					
ORGANISM	Homo sapiens.				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE					
AUTHORS	Seifert,M., Engel,M., and Welter,C.				
JOURNAL	Intersectin 2 (SH3D1B), human homolog of mouse Ees2 protein				
REFERENCE	2 (bases 1 to 5828)				
AUTHORS	Seifert,M., Engel,M., and Welter,C.				
TITLE	Direct Submission				
JOURNAL	Submitted (23-MAR-2000) University of Saarland, Institute of Human				
	Genetics, Building 68, Homburg 66421, Germany				
FEATURES	Location/Qualifiers				
source	1..5828				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/chromosome="2"				
	/map="2p"				
	/cell_line="HEL"				
	1..5828				
	/gene="SH3D1B"				
	7..5052				
	/note="adaptor protein; similar to Mus musculus ESE2				
	protein"				
	/codon_start=1				

```

/product="intersectin 2"
/protein_id="AAF63600.1"
/db_xref="GI:7542783"
/translation="MASEYQKLRPTMAOFTAMNGGPNMKAITSEERTKHDFQDFNL
KPSGGYITVDQARNFPLDSGLPAPVLAETMALSLNKGKDDQDFEIAKLILKLIQ
GQGLPEVLPIMKQPMFSPILISAFMGSPMNLSTIQPLPAAPITLSLSATISNL
PLMLPPTLVPSVSTSLPNGTASLIQPLIPYSSSTLPHSSSYSLMGSGASIQK
AQLSLDSSSTSTASLNGSPKTSSEMAVPOPTRLKROCTNLDKMSGYLSIQK
FOARNAILOSNIQOTLATIWTLDVDDGQLKEEFLIHLNHLIDMAKAGQPLPLP
PELVPSFRGKQIDISNGTLPSYOKMOEEEPKKLPTFEDKIDKAVENGMLEKAR
RQALMEQOQAEBAKQKKEKEERKQELQDEQKQLEIKLEKQELEREREE
RRKDIEREAQKQELERERLEWERIRQELNKNQKQEBETVLNKKKLIHLA
LNKHQOISGRLOVRLKQKQTELEVDKQCOLEIMEIQLOQDEKQONKLIYLV
PEKOLINERIKNQFNSPTDSGSLHKKLEKEELCORLEQDQDALEKEDASLMS
DSFNQOLKEIETVNTQALBOLYKIKRDLKEIRLEKLEKLEKLEKLEKLEK
GKENIMKLEKEERKQKLEKEERKQKLEKEERKQKLEKEERKQKLEKEERKQ
SVLVNRYALYPERARNDHMSFNSGDIQVDEKTVGEFGMLYSGFQNGFPCPNYB
KMPSENEKAVSPKXALPPTVLSLSTISSEPSLNOBPASVTOYVNSFNSLTVNS
WOKSAFTRTVSPGSVPDIHQGVENVNKAALCMTAKDNHNLNFSKDIITVLQO
QENMFGEHGGRGWPKSYKIIPGSEVKEEPEALYAAVNPVVKDQESGSA
LYPSVSRGDLTFEGBEILVTKQDGMWMTGSIQDRGIPSNVYVKDQESGSA
KSGANRKEPEIAQVTSAYVAGSGSRLSIAPGOLLIIKKNISGMWQELARQKRRK
GMPASHVYKLGPSSEERATPAHPYCOIYAMVDYANNEDLSKQOLIVNMKDDP
DMWQELNGVTLGPPSNVVKMTTSDPSQWQCADQLDTPQLEKROQYIHLITQ
BERYADQLVVEVFOKMASSGLTBEGMALIFVNMKELIMSTKLKALRYKKTG
GEMPVOMIGILIAELSHMOAYIRFCQNLGALLOQKEDDEDFEFLKIASDP
RCKMPSSFLPKPMQIRITRYPLIRSLILENTPSHADHSLKLALBAEELQVNE
GVREKENDRLKEMIOAHVOCGLAEOLFNSLTNGIPRKLHSGKLYKTKNSNELLG
FLNDFLLITVWYKQFAVSSSEKLFSSKSNAGKRMKTPFLNEVLYKLTDDSSDE
PVFHLSHIDRYVTLKTDNIINERTAWQKIKASBOYIDTERKQEKAVQASQTSGL
GLYOVHLVATELKAQKPNKSNPYCELSMGSOSYTTTITDITLNPKNFPCQPIID
LYOVLCITLFPDRQSPDPLGRPEIIVAKIRTEQSKGMTRRLILHEVPTGEVWY
RFDQLFPOKTLT"

```

BASE COUNT 1867 a 1233 c 1337 g 1391 t  
ORIGIN

Query Match 19.8%; Score 721; DB 9; Length 5828;  
Best Local Similarity 53.9%; Pred. No. 1.9e-152;  
Matches 1998; Conservative 0; Mismatches 1430; Indels 279; Gaps 14;

```

QY 1 ATGGCTCAAGTTTCCACACCTTTGGTGTGAGTCTGGGCCATTAAGTGGAG 60
DB 43 ATGGCTCAAGTTTCCACACCTTTGGTGTGAGTCTGGGCCATTAAGTGGAG 102
QY 61 GAAAGGCCAAGATGACGAGCTCTTAAAGCCATAGCGGATTTAATTACT 120
DB 103 GAAAGGCCAAGATGACGAGCTCTTAAAGCCATAGCGGATTTAATTACT 162
QY 121 GGTGATCAAGCGAAGAACTTTTTCATCTGGGTACTCAGCTGTCTTAGACAA 180
DB 163 GGTGATCAAGCGAAGAACTTTTTCATCTGGGTACTCAGCTGTCTTAGACAA 222
QY 181 ATATGCGCGTACCGGACATGAATTAACGATGAGATGATGAATTTTTCATA 240
DB 223 ATATGCGCGTACCGGACATGAATTAACGATGAGATGATGAATTTTTCATA 282
QY 241 GCGATGAAGCTTCAAACTGAAGCTACAGATATCAGCTCCCTCCACACTTCCCT 300
DB 283 GCGATGAAGCTTCAAACTGAAGCTTCAAGGCAACAGTTGCTGTCTTCCCTCT 342
QY 301 GTCATGAACAGCAACAGTGG---GCTATTTCCAGTGCACACAGATTTGGATAGAGG 357
DB 343 ATTATGAAGCAACCCCTTAAGTTTCTCATTAATTTGCTGTGTTTGGAAATGGGAAC 402
QY 358 ATTGCTAGCATG-----CCACCACTCACAGTGTGCT 390
DB 403 ATGCCCATCTGTCCATCTCTCAGCAATTCCTCAGCTGCACCTTAACATCATTTGCT 462
QY 391 CCGTGCCTAATGGGCTCA---TTCCAGTTGTTGAATGCTCCACCTTAAGTATCTTC 447
DB 463 TCTGCGACTTCAAGGACCAACTTCTCTCTTAATGATGCGCCCTCCCTAAGTGCCTTCT 522
QY 448 GTCCCTTCAGACAGTGTCTCCCTGCTTACGAGGCTTCTCTCCGTCAATACAGCTCTG 507

```

```

DB 523 GTTAGCACATCATCATTAATCCAAATGGAACCGCAGCTCATTTAGGCTTTACCATTTCT 582
QY 508 CCGGCGTTGGCATCCTGACGACATAGGCGCAAGAGTCTCTTTCAGACATCTGCT 567
DB 583 TATCTTCTTCAACATTTGCTCTCATGGGTATCTTTATAGTGTGATGGAGATTTTGA 642
QY 568 CCAAGGTACAAATTAACACTAAGTTACAGAAAGCAATCATTCAGATGTGGCAGCGCC 627
DB 643 GGTGCTAGTATACAGAAAGGAGCTCTGATTTGATTTAGATCTAGTCAACTTCC 702
QY 628 CTTCCAG-----CAGCAAAATGGGCTGTGCTT 654
DB 703 TCGACTCTTCACTCTCAGGAACTCACCCAGACTGGGACTCAAGTGGGAGTTCCT 762
QY 655 CAGTCATCAAGGCTGAAATACAGGAGTTATTCACAGCCACAGCAAAATATAGATGGA 714
DB 763 CAGCTTCAAGATTAATAATATCGGCAAAATTTAATATCTTACCAAAATATAGATGGA 822
QY 715 CACTTAACAGTCCCGCAGGACAAATATTCATGCAATCAAGTTTACCCAGGCTCAG 774
DB 823 TATCTCAGGTTTTCAGAGTAAAGTATGCTTCTTCAATCATTCATTCATCAATCTCAG 882
QY 775 CTGGCTTCAATATGGAATCTTTTGTGACATTGATCAAGATGGAATCTCATCGAAGAA 834
DB 883 CTGGCTCTATTTGGACTCTGCTGAGCTGATGATGAGACAGCTTAAAGCAGAAAG 942
QY 835 TTTATCTAGTATGACCTAATGATGTTGCAATGCTGTGAGCCAGCCAGCCGCTG 894
DB 895 CTGCTCAGAAATACCTCTCTCTCTTCTTCAAGAAAGTTGCTCTCGGACGTGGAGTCC 954
QY 1003 TTACTCTCAGTGTGTTTCTCTTCTTCTTCTTCAAGAGAAAGCAAAATGATTCATTAAT 1062
DB 955 GTCAATTAAGCTTTCTTCTGTGATCAAGGCTGCTGAGAGCCGTGTGAGAGATGAG 1014
QY 1063 GGAATCTGCTTCTATATCAAGAA-----AATGCAAGAA 1095
DB 1015 CAGAGCAGAGAAAGAAATGCGCTGTGACATTTGAATGAAGAGCGGAGAACTTGCAG 1074
DB 1096 GAGAGCTCTCAAGAAATTAACAGTTACTTTTGAAGACAAAGGAAAGCCACTATGAG 1155
QY 1075 CGAGGACAGTGTGAGCTGTGAAACCGCCGCAAGCGCTTTGAGCAGCAGCGCAAGAG 1134
DB 1156 CGAGGACAGTGTGAGCTGTGAAACCGCCGCAAGCGCTTTGAGCAGCAGCGCAAGAG 1215
QY 1135 CAGAGCGGTGCTCAGCTGAGAGCGCGGAGCAGAGAGAGAAAGCGGAGCGCCAG 1194
DB 1216 GCAGAACGTAAGGCCAGAAAGAAAGAAAGAGAGAGAGAGAGAAAGAGAAATTA 1275
QY 1195 GAGCAGAGCGCCAGAGCGGAGCTGAGCTGGAAGAGAGCTGGAAGAGCAGCGGAGCTG 1254
DB 1276 GAAACAAGATGAAGAAACAACTTGAATTGAAAAACGCTTAGAGAGCAACGGGAATG 1335
QY 1255 GAGCGCAGCAGAGAGAGAGAGAGAGAGAGATTCAGAGAGCGCGAGCGCAAAACGG 1314
DB 1336 GAGAGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1395
QY 1315 GAACTGGAAGAGCAGCGCAAACTTGAATGGGAACGAAACCGGAGACAGAACTCTGAAT 1374
DB 1396 GAACTTGAAGAGCAGCGCTTGAATGGGAAGAGAAATTCGGGACAGGAGCTTCTCAAT 1455
QY 1375 CAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1434
DB 1456 CAAAGAGATTAAGAAACAGAGAAATTTGTCAAGTTAACTTTAAAAAGAGAACTTCTCAT 1515
QY 1435 TTTGATTAAGAGCTCTGAATGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGATATC 1494
DB 1516 CTTGATTTGAAGAGCTGAATGAGCAAAATCAAGCAGATTCAGGACAGATTCAGAGATGC 1575
QY 1495 AGGTGCACTGAGCAACCCAGAGGAGAGAAATTAAGAGCAGAAAGTCTAAGAGACTTA 1554

```

Db 1576 CGACTCAAAAAGCAAACTCAAAAAGACTGAGCTGGAAGTTCTGGATAAGCAGTGTGACTTG 1635  
 QY 1555 AGAATGCTGAATACACCCCTTAACGACGACAGATTGGAGAAATCTCAGCAAAAGCTTGA 1614  
 Db 1636 GAAATTAATGGAATCAAGCACTTCAACAGAACTTGAAGAAATACAGAAATGACTTATC 1695  
 QY 1615 AGACTTAATTCAGAGAAACAGATATCTCAGTGCAGATTAAACAGTTCAGCAGAAACAGT 1674  
 Db 1696 TATCTGGTACTGAGAAACATATTAATTAAGAAATTAATAAATCATGCGAGTTACAGTAAC 1755  
 QY 1675 TTGCTATGAACTCGCTTCTTACCCTCAAAAGACCTTGAAGCAAGAGAGAGCTGGCCGG 1734  
 Db 1756 ACACCTGATTCAAGGGGTCAGTTACTTATTAATAAATCATTAAGAAAGGAATTAATGC 1815  
 QY 1735 CAGCAGCTCCGGAGCAGCTGGAAGAGTGGAGAGAGAGACAGAGTCAAAAGCTGACGAG 1794  
 Db 1816 CAAGAATTAAGAAACGTTAAGATGCTTGAAGAAATTCATCACTAAGCTGTAGAA 1875  
 QY 1795 ATTGATTTTTCACCAACCAAGCTGAGAGAACTGAGAGATATATGCAAAACGCAACTC 1854  
 Db 1876 ATGGATTTCTTTAACTCAATCACTAAAGAACTGAGAGAACTCAACACACGCACTTA 1935  
 QY 1855 CAGAGCAGAGGTCCTCGAGGAGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1914  
 Db 1936 GCCTTGAACAGCTTTATAGATCAAACTGACAAAGTTGAAGAAATTTGAAGAAAGAA 1995  
 QY 1915 CTGAGTTAGAGAAAGCAAAAGGAAGACGCTCAGAGACGAGTTGAGAAAGGCAAGCAA 1974  
 Db 1996 TTGAACTAATGAGAAAGAAAGAACTAG-----AAGATG 2029  
 QY 1975 TTGCTGAGCATGTGCGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2034  
 Db 2030 AGGCTGCAAGAAAGCAAG 2089  
 QY 2035 AGACTGAAGAGGAG 2094  
 Db 2090 AGGAAAG 2149  
 QY 2095 ATGCAAG 2154  
 Db 2150 AGGAAAG 2202  
 QY 2155 CAGGACCTGCTGCTACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2214  
 Db 2203 GAGGAG 2230  
 QY 2215 AAGTGTATATTAACGAGCGCTGTACCCCTTTGATCCAGAAATCAAGATGAGATCACC 2274  
 Db 2231 -TTTTGTGATTAATAGAGATTAATCCCTTTGAGAGAGAGAGAGAGAGAGAGAGAGAG 2289  
 QY 2275 ATCCAGCAGAGAGATATAGTATGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2334  
 Db 2290 TTTAATTTCTGAGATTAATTAATTCAGATTGATGAGAGAGAGAGAGAGAGAGAGAGAG 2349  
 QY 2335 GAG 2394  
 Db 2350 TATGATGATTTTCAAG 2409  
 QY 2395 GAAATATGAGTTCCCACTCCAGCCAAACAGTGCAGATGAGATGAGAGAGAGAGAGAG 2454  
 Db 2410 TCAAGTGAATAAT-----GAAAAAGCTGTATCTCAAGAGAGAGAGAGAGAGAGAG 2460  
 QY 2455 AAGTGTGCTGCTGAG 2514  
 Db 2461 AAGTGTGCTGCTGAG 2520  
 QY 2515 ACCCCCAACAGTGGAG 2574  
 Db 2521 TCAAGTGTGATTAATCAAAA----- 2540  
 QY 2575 GAAAG 2634  
 Db 2541 -----TGATCTTTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2571

QY 2635 CAGTTAGGAGAGAGATCAGCTTTTACCCAGCCACAGCAGAGAGAGAGAGAGAGAGAGAG 2694  
 Db 2572 TCATGGCAG 2628  
 QY 2635 GTCTGGGCGAGGAG 2754  
 Db 2629 ATTATGAG 2688  
 QY 2755 GCGAAAAAG 2814  
 Db 2689 GCAAG 2748  
 QY 2815 CAAGCATGTGTGATTTGAG 2874  
 Db 2749 CAAG 2808  
 QY 2875 GTGAAGCTCATTTCAG 2934  
 Db 2809 GTCAAGATCATTTCTGGAG 2856  
 QY 2935 AGTCTGCTAGTCTAAG 2994  
 Db 2857 TTGATGAG 2910  
 QY 2995 GAGTTTATGCTATGATACATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3054  
 Db 2911 GAAATATATGACCTTTTATCAATTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2970  
 QY 3055 GGGAGATGATTTGATGATACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3114  
 Db 2971 GGTGAAG 3030  
 QY 3115 AAGTCCGAGGCTTCTTCTTATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3174  
 Db 3031 AGAAGTGAATTTTTCATCACTATGATCAACCAAGAGAGAGAGAGAGAGAGAGAGAG 3090  
 QY 3175 GGTGGAG 3234  
 Db 3091 GCTAGCAG 3150  
 QY 3235 GCTGCTATGCTCCGAG 3294  
 Db 3151 GTTGCTTGTGCTTGAAG 3210  
 QY 3295 AAG 3354  
 Db 3211 AAAAAATCAAGTGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3270  
 QY 3355 GGTGCTTTCCAG 3414  
 Db 3271 GGATGTTTCTGCGAG 3330  
 QY 3415 ACTGAGCTACCAAG 3474  
 Db 3331 GCTTTATATC-----CTGTATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3369  
 QY 3475 TACAGCGCCAG 3534  
 Db 3370 TATGAG 3429  
 QY 3595 TCCATTTATGTAAGCTGAG 3641  
 Db 3490 TCAACTATGTTAAGATGAG 3536

RESULT 14  
 AB033082  
 LOCUS AB033082 5938 bp mRNA linear PRI 11-NOV-1999

DEFINITION Homo sapiens mRNA for KIAA1256 protein, partial cds.  
 ACCESSION AB033082  
 VERSION AB033082.1 GI:6330910  
 KEYWORDS  
 SOURCE Homo sapiens brain cDNA to mRNA, clone\_11b; pBluescriptII SK plus  
 clone:h15293.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (sites)  
 Nagase, T., Ishikawa, K., Kikuno, R., Hirose, M., Nomura, N. and  
 Ohara, O.  
 Prediction of the coding sequences of unidentified human genes. XV.  
 The complete sequences of 100 new cDNA clones from brain which code  
 for large proteins in vitro  
 DNA Res. 6 (5), 337-345 (1999)  
 JOURNAL 2 (bases 1 to 5938)  
 MEDLINE Ohara, O., Nagase, T. and Kikuno, R.  
 REFERENCE Direct Submission  
 AUTHORS Submitted (04-OCT-1999) Osamu Ohara, Kazusa DNA Research Institute,  
 Laboratory of DNA Technology; 1532-3 Yama, Kisarazu, Chiba  
 292-0812, Japan (E-mail: cdnainfo@kazusa.or.jp,  
 URL: http://www.kazusa.or.jp/huge/, Tel: +81-438-52-3913,  
 Fax: +81-438-52-3914)  
 FEATURES  
 source Location/Qualifiers  
 1..5938  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="h15293"  
 /tissue\_type="brain"  
 /clone\_11b="pBluescriptII SK plus"  
 1..5938  
 /gene="KIAA1256"  
 <148..5178  
 /gene="KIAA1256"  
 /note="Start codon is not identified."  
 /codon\_start=1  
 /product="KIAA1256 protein"  
 /protein\_id="BA06570.1"  
 /translation="OgkLRRTMAAOPFTANNGSPNNMALTSEERTKHDFQNLKPSGG  
 /db\_xref="GI:6330911"  
 /db\_xref="Xref:6330911"  
 /translation="OgkLRRTMAAOPFTANNGSPNNMALTSEERTKHDFQNLKPSGG  
 VILPDMKQPMSPPLISARFGSMSPNLSIPPLPAPRPLISLSSATSGTLLPLM  
 PPLVSVSTSSIPNGTASLIQPLPIPIYSSSTLPHSSSTSLMGFGGASIOKASLI  
 DLGSSSTSTASLNGSPKRTSEMAVPOPLRLKQKQFNTLDSMSGYLSGFGARN  
 ALIOSNLSTQTLATITLADVDGDLKAEFLAMHLIDMAKAQPLPLTLPPLV  
 PSFSGKOIDSINGTLPSYQKQDEEPQKLPVTFEDKKAVERGNMELERQALM  
 EOOQREARRKAKOEKEMERKOELOEOKKOLEKLEKRELORELOREERERKDI  
 ERREAKOLERORLEMERIRQELINKRDOEIVLNNKKXKHLLEALNGKH  
 OQISGRLOVRLKQOTKTELEVLDKQCLELMEITKQDOLEOBYONKLIIVPERKOL  
 LNERIKNMFNTSDSVSLHKKSLKEKLEQRLKEQDLDALEKTAASKLSDMFNN  
 OUKELREYNTQALAEOLYKIKRDKLEIKERRLELMOKKLEBAARKAKQGENL  
 MKNLEKEEKKOKLEKTOEKIOEERKAEKOKKDKTLKAEKRETAAYLVN  
 YRALYFEARNHDEMSGNDIIQVEKTVGEGMLYSGFQNGFNGFNVNEMKPS  
 ENKKAASPKKALLPTVLSATSTSEPSLNQPSAATVQNVVSNLTVMNKKSS  
 ATRTVSPGSVPTFGQGVENLKAQALCSWTAKDNHLSKDIITVLEOENMM  
 FGEVHGKGMFPKSVYKIIIPGSEVKEEPEALYAANKRPTSAASVGEVIALPYIS  
 SVEPDGLTFEGEELIVTQKDEMTGSGIDRSGLIPSNVYVRKQDESGASKSGAS  
 NKPEIAQVTSAAVAGSEQLSLAQQLILIKNTSGMWQELQARGGKROKGFPA  
 SHYKILGPSEERATPAFHVCOVILAMDYAANNEDLSFSGQLINVMKDDPDMWG  
 EINGVGLFSPSVKMTTDSQOCADLOTLDMOPIERRGOGVHELIIOTEEYRM  
 ADIOLVVEVFOKMAESGFLTBGEMALIVNNKELINSTTKLIKALPRKKTGGKMP  
 VOMIGDILAEISHMOAYIRPSCQNLGALALQOKDETDREPLFKLASPRCKGM  
 PLSSFLKPMORITRYPPLIRSLIENTPESHADSHSLKALRAEELCQVNEGKEX  
 ENDRLEMIQAHOCCEGLAEOLIFSLTICIGRLKHLGKLYKTSNKEHLGFLFND  
 FILITVMVOFAVSGSEKLFSSKNAOFMYTPIELNEVLYKPTDSSDEPEHI  
 SHIDRYVTLPTDINERTAMVOKIKASNAOFMYTPIELNEVLYKPTDSSDEPEHI  
 HUIEATELKACKNGKSNPYCEISMSGOSTTTTIDTLPNKNPCQFIIDYIDYV  
 LCILFLDRDOFSPDPLGRTEIPVAKIRIODESKGPMTRRLLEHVEGTGAVWRPDLQ  
 LFEOKTLL"

Query Match 19.8%; Score 721; DB 9; Length 5938;  
 Best Local Similarity 53.9%; Pred. No. 2e-152;  
 Matches 1998; Conservative 0; Mismatches 1430; Indels 279; Gaps 14;  
 1 ATGGCTAGTTTCCCAACCTTTGGTGTAGAGCTGATGTGGGCAATCTGTGAG 60  
 169 ATGGCTAGTTTCCCAACCTTTGGTGTAGAGCTGATGTGGGCAATCTGTGAG 228  
 61 GAAGGGCAAGCAATGACCAAGCTTCTAGCCTGAAGCCATAGGGGATTTACT 120  
 229 GAAGGGCAAGCAATGACCAAGCTTCTAGCCTGAAGCCATAGGGGATTTACT 288  
 121 GATGATCAAGCAAGCAATTTTCTTCCAACTTGGTGTAGAGCTGATGTGGG 180  
 289 GATGATCAAGCAAGCAATTTTCTTCCAACTTGGTGTAGAGCTGATGTGGG 348  
 181 ATATGGGGCTAGGGGACATGAATGAATGAATGAATGAATGAATGAATGAAT 240  
 349 ATATGGGGCTAGGGGACATGAATGAATGAATGAATGAATGAATGAATGAAT 408  
 241 GCAATGAAGCTTATCAATGAATGAATGAATGAATGAATGAATGAATGAAT 300  
 409 GCAATGAAGCTTATCAATGAATGAATGAATGAATGAATGAATGAATGAAT 468  
 301 GTCATGAAGCAAGCAAGCAAGT---GCTATTTCAGTCAAGCAAGCAAGT 357  
 469 ATATGAAGCAAGCAAGCAAGT---GCTATTTCAGTCAAGCAAGCAAGT 528  
 358 ATATGAAGCAAGCAAGCAAGT---GCTATTTCAGTCAAGCAAGCAAGT 390  
 529 ATATGAAGCAAGCAAGCAAGT---GCTATTTCAGTCAAGCAAGCAAGT 588  
 391 CCGTGGCAATGGGCTCA---TTCCAGTTTGGAAATGTCACACCTTATCTTCT 447  
 589 TCTGGACCTTCAAGGACCAACCTTCTCTTAAATGAAGCCACCTTCTCTCT 648  
 448 GTCCTCAGCAGCAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 507  
 649 GTTACACATCATCTATTACCAATGAAGCAAGCAAGCAAGCAAGCAAGCA 708  
 508 CCGTGGTTCGATCTCTGACCAATGAGCAAGCAAGCAAGCAAGCAAGCA 567  
 709 TATCTCTTCAATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 768  
 568 CAGGCTCAATTAATCAATGAATGAATGAATGAATGAATGAATGAATGA 627  
 769 GGTGCTAGTATCAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 828  
 628 CCGCAG-----CAGCAGAAATGGGCTGTGCT 654  
 829 TCGACTCTTCACTCTGAGGAATCAAGCAAGCAAGCAAGCAAGCAAGCA 888  
 655 CAGTCAATGAGCTGAAATGAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 714  
 889 CAGCTCAATGAGCTGAAATGAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 948  
 715 CACTTAACAGTCTCCAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 774  
 949 TATCTCTAGGTTTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1008  
 775 CCGGCTCAATGAGCTGAAATGAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 834  
 1009 CCGGCTCAATGAGCTGAAATGAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1068  
 835 TTATCTAGTATGAGCTGAAATGAAGCAAGCAAGCAAGCAAGCAAGCAAG 894  
 1069 TTATCTAGTATGAGCTGAAATGAAGCAAGCAAGCAAGCAAGCAAGCAAG 1128  
 895 CCGCTCAAGTATGAGCTGAAATGAAGCAAGCAAGCAAGCAAGCAAGCAAG 954  
 1129 TTATCTAGTATGAGCTGAAATGAAGCAAGCAAGCAAGCAAGCAAGCAAG 1188

QY	995	GTATTAAGCTCTTCTTCGTGTGATACAGAGGCTGCTGTAGAGAGCCGTCGTACAGAGGTGAG	1014
Db	1189	GAATCTCTGCTTCTCATATCAGAA-----AATCCAGAA	1221
QY	1015	CAGCAGCCAGAGAAAGAACTGCTGTGACATTTGAAAGTAAAGCGGAGAACTTTCGAG	1074
Db	1222	GAGGAGCTCTAGAAAGAAATTACAGTTACTTTTATGAGCAAAACGAAAGCAACTATAG	1281
QY	1075	CGAGGCACTGTGTGAGCTTGGAGAAACGCGCCGACAGCGTCTTTGGAGCAGCAGGCAAGAG	1133
Db	1282	CGAGGGAACATGTGAGCTGGAAAACGACGCCAGCTTGTATGAGAGCAGAACAAAGGAG	1341
QY	1135	CAGGAGCCGTTTGCTCAGCTGTGAGCGCGGCGAGCAGAGAGGAAAGAGCGGAGCGCAG	1194
Db	1342	GCAGAAACGTAAAGCCCAAGAAAGAAAGAGAGTGGGAAAGAAACAGAGAGATTATCA	1401
QY	1195	GAGCAGAGGAGCCAGCGGCACTGTGAGTGTGAGAGCAGCTGTGAGAAACAGCGGAGCTG	1254
Db	1402	GAACAAAGATGAGAAACAACTTGAAATTAGMAAAACGCTTAAAGAGACAGCGGAGATTG	1461
QY	1255	GAGCGGACAGAGAGAGAGAGAGAGAGAGATTCAGAGGCGCGCAGGCGCCAAACCG	1314
Db	1462	GAGAGACACACGAGAGAGAGAGAGAGAGAGACATAGAAAGAGAGAGGCACTAAACG	1521
QY	1315	GAATCTGAAAAGCAGCGCAACTTGAAATGCGAAACGGAACCGAGACAGAACTCCTGAAT	1374
Db	1522	GAATCTTGAACCAACACGTGCTTGAAATGGAGAGAAATTCGCGCAGACGAGACGTTCTCAAT	1581
QY	1375	CAGAGAAACAGAGAGCAGAGAGGACCGTGGTCTCTGAAGCAAGAGAGAAAGCTCGGAG	1434
Db	1582	CAAAAGAAATAGAGAACAGAGAAATTTGTCAAGTTAACTTTAAAGAAAGATCTTCAAT	1641
QY	1435	TTTGAGTTAGAACTCTGAATGACAAAGAACATCAGCTAGAGAGAAACTTCAGATATC	1494
Db	1642	CTTGAGTTGAGAGCACTGAATGTGCACAAACATCAGACAGATCTCAGCGAGACTTCAGAGATGC	1701
QY	1495	AGGTGTCAGCTGTGCACCCAGAGCAAGAAATTGAGAGCAGCAAGCTGTAGAGACTA	1554
Db	1702	CGACTCAAAAAGCAAACTCAAAAACCTGAGCTGAGAGTTCTGTATAGAGAGGTGATCTTG	1761
QY	1555	AGAAATTCGTGAATACCCACTTCACAGCAGCAGTGTGAGAGATCTCAGCAAAATCGTTGGA	1614
Db	1762	GAAATTATGAGAAATCAGAGCACTTCAACAGAGAACTTCAGGAATATAGAAATAGCTTATC	1821
QY	1615	AGACTTATTTCCAGAGAAACAGATCTCAGTGACCACTTAAACAGTCCAGCAGAAACGT	1674
Db	1822	TATCTGCTACTCTGAGAGCAATTTTAAATGAAAGAAATTTAAACATGTGACGTTTCAGTAAC	1881
QY	1675	TTGCAATAGAGACTCGCTTCTTAACCCCTCAAAAGAGCCTTGGAACCAAGAGAGCTGGCCGG	1734
Db	1882	ACACTGATTTACGGGGTCACTTACTTCTCATTAATAAATCATTTAGAAAAGAGAAATTAATGC	1941
QY	1735	CAGCAGCTTCGAGAGACAGCTGTGACGAGGTGAGAGAGAGACGAGCTCAAGCTGCAGAG	1794
Db	1942	CAAAAGACTTAAAGAACAGTTAGATGCTCTTGAAAAAGAAATTCATCTAAGCTGTCAAGAA	2001
QY	1795	ATTGATGTTTTCAACACACAGCTGTAAAGAACTAGAGAGATATATGCAACAGCAATCC	1854
Db	2002	ATTGATTTCTTTTAAACATCACTAABGAACTGTAGAAACCTCAACAACACAGAGATTTA	2061
QY	1855	CAGAAACAGAGGTCCCTGTGAGAGCCGCGACTGAAGCAGAAAGACGAGAGAGAGAAAGAC	1914
Db	2062	GCCCTTGAACAGCTTTTATTAAGATCAAAAGTGAACAGTTGAGAAATTCAGAGAAAGAA	2121
QY	1915	CTGAGATTAGAGAAACAAAGAGACGCTCAGAGACGAGTTCAAGAAAGGAGCAAGCA	1974
Db	2122	TTAGAACTTAATGCAAGAAAAAAGAACTAG-----AAGATG	2155
QY	1975	TGGCTGTGAGCATGTGCGACAGAGAGAGACACACGCCCCCGGAAACCCACAGAGAGAC	2034
Db	2156	AGGCTGTGAGAGAAAGCAAGACAGAGAAAGAAACCTTATGAGAAAGAAATCTTTAGAAAG	2215
QY	2035	AGACTGAAGAGGAAAGACAGTTGTGAGAGAGAGAGCGGAGAGAGACCAAGCCGAGAA	2094

Db	2216	AGGAAGAAAGAAAA	CAAAACGCACTCCAGAGAA	AAAAAACA	CAAGAAAAAATTCAGAGAG	2275
Oy	2095	ATCGAAGCAAGCAGAGT	CGGCTTTTCCATCCGCATCAGAGCCAGCTA	AGCTGCGCAC	2154	
Db	2276	AGGAACGGAAAGCTGAGGAGA	-----	AACAACGTAAAGATTAAGATATCTTTGAAAGCT	2328	
Oy	2155	CAGGACCCCTGCTACCA	CAGAGAAAGGCCGCTTACATTTCTGCA	CAGAGAGGTGA	2214	
Db	2329	GAGGAGAAAAAA	CGTAGACAGCTAGTG	-----	2356	
Oy	2215	AAAGTGATATATTA	CCGAGCGCTGACCCCTTTGAATCCAGAA	GTACAGATGATGAC	2274	
Db	2357	-TTTGTGGAATTA	TAGAGCATTAATCCCTTTGAAGCAAGAAC	CCATGATGATGATGAT	2415	
Oy	2275	ATCCAGCCAGAGATAT	AGCATATGATGTGAATGAAGCCACATGAGAGCCAGAT	AGGCTT	2334	
Db	2416	TTTAATCTGGAGATAT	TAATTCAGTTGTATGAAAAAA	CCGTAGGAGAAC	CTGTGTGGCTT	2475
Oy	2315	GAGAGAGAGCTGAAG	AGGAGAGCGGATGTTCCCTGCAAACTATATG	CAGAAAAAATTTCCA	2394	
Db	2476	TATGTATGTTTTCA	AGGAAATTTTGGCTGTTTTCCATGCAATTA	TATGTGAAAAAATGCCA	2535	
Oy	2395	GAAATATGAGTT	CCCACTCCAGCCAAACAGATGCCATCTGACATCTG	CCCTGCCCC	2454	
Db	2536	TCAAGTGAAT	-----	GAAAAAGCTGATATCCAAAGAGGCTTATCTTCTCT	2586	
Oy	2455	AAACTGCTCTGCGTGAG	ACCCCTGCTCTCTTGGCCAGTACCTCTTCTGAGCCCTCCACA	2514		
Db	2587	ACAATTTCTTATCTG	TACTCAATTTCTCTGAAACCACTTTCTTCAATATCA	CCAGCA	2616	
Oy	2515	ACCCCAACAATCGG	CAGACTTCACTTCCAGTGCC	CCAGAGCTCAAA	CGAGAGCCA	2574
Db	2647	TCAGTGACTGATTA	TCAAA	-----	2666	
Oy	2575	GAAACGACAACTGG	AGATAGTGGGGCGTCAGCCCTTCTGACCGTAC	TAGTGCTGC	2634	
Db	2667	-----	TGATCTTTTCAAACTTA	CTGTATAATCA	2697	
Oy	2635	CAGTTACGGAGAGAT	AGCTTTTACCAGCACAAGCCACATGCTCTCCCATCTTCC	2694		
Db	2698	TCATGGCAGAAAAAT	TACGCTTCACTCGAAC	TGATGTCCTCGG--	ATCTGATACACT	2754
Oy	2655	GTCCTGGCCAGGGT	GAAAAAGGTGGAAGGGCTTAAAGCCAGACCCCTGTATCC	CTTGAGA	2754	
Db	2755	ATTATGACAGGAG	CAAGTGAGAAAACTTAAAGCAAGCCCTTGTCTTGAGCT	2814		
Oy	2755	GCCAAAAAAGAC	CAACTTAATTTTAA	CAAAAGTGAGCTATCACGTTCTGGA	CAG	2814
Db	2815	GCAAGAAAGATTA	CACTTGAACCTTCAAAACATGACATTA	TATCTGTCTTGAGCAG	2874	
Oy	2815	CAAGACATGTGG	GTGTTTGGAGAGTTCAAGTCA	AGAGGGTTGTTCCCAAGTCTTAC	2874	
Db	2875	CAAAAAAATTTGG	GTGGTTTGGGGAGTGACATGAGAGAAAGATGTTTCC	CAAAATCTTAT	2934	
Oy	2875	GTAAGACATCATTT	CAGGACCCGTAAAGAAATCCA	CAGATGATATGAGCCCTACTGAA	2934	
Db	2935	GTCAGATCATTT	CTGAGATGAAGTAAACGGAA	-----	GAACCAAGAGCT	2982
Oy	2935	AGTCTGCTAGT	CTAAAGAGATGGCTTCCCGGCGCCCAAGCC	AGCAATTC	CCGGAGAA	2994
Db	2983	TTGTATCAGCTGAT	TAATAAAGAACTTACCTCGGACGACTTATTCAG	-----	TTGGAGAA	3036
Oy	2995	GAGTTTATTC	CCATGACATACGAGACTTCTGAGCAAGAGATTTATAC	CTTTACAGAA	3054	
Db	3037	GAAATATATG	ACCTTATTCATATTCATAGTGTGGAACCTGAGATTTA	CTTACACAGAA	3096	
Oy	3055	GGGGATCTGAT	TGTGTGTTACAAAGAAAGATGTGATGTTGAGCCGGAA	ACGTTGGCGAC	3114	
Db	3097	GGTGAAGAAATAT	TTGTGTGACCCGAAAAAGATGTGAGTGAGTGTGACAGAA	GTATTTGAGAT	3156	
Oy	3115	AAGTCCGAGT	CTTCCCTTCTAACTATGTGAGGCTTTAAAGATTCAGAGG	CTTGGAACT	3174	



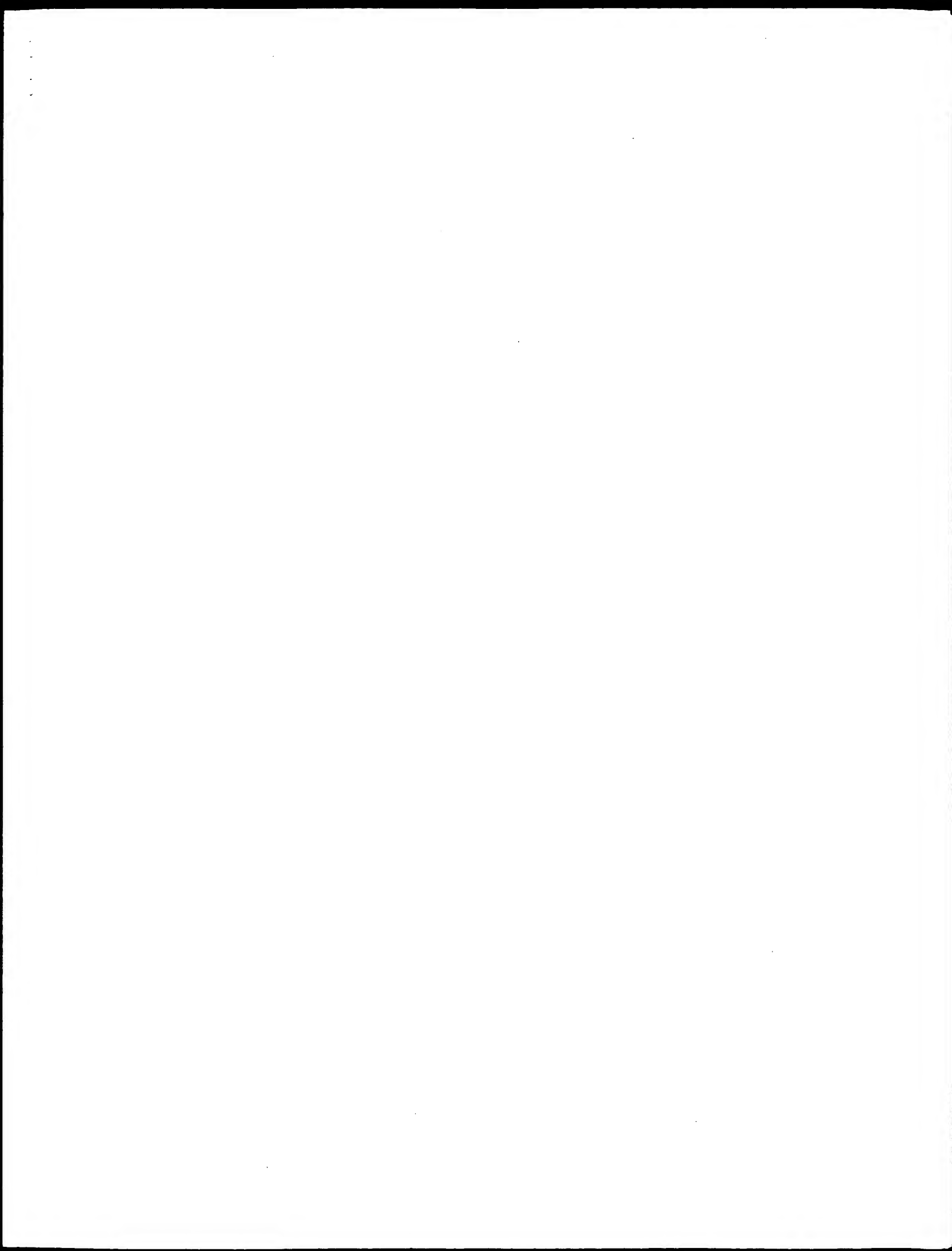


Db 841 GGTGCTAGTATACAGAAAGCGCAGTCTCTGATTGATTAGATCTAGTACTCAACTTCC 900  
Qy 628 CTTCCAG-----CACCAATATGGGCTGCTCCT 654  
Db 901 TCGACTGCTTCACTTCAGGAACTCACCAGACTGGACCTCAGAGTGGGCACTTCTCCT 960  
Qy 655 CAGTCATCAAGGCTGAATACAGGCACTTATTCACAGCCACGCAAACTATAGTGA 714  
Db 961 CAGCTTACAAAGATTAAATATCGGCAAAATTTAATATCTTGAACAAAGTATGAGTGA 1020  
Qy 715 CACTTAAAGSTGCCAGGCAAACTATTTCTCATGCAATCAAGTTTACCCAGGCTCAG 774  
Db 1021 TATCTCTAGGTTTTCAGCTTGAAGATGCCCTTCTTCAAGTCAAACTTTTCTCAACTCAG 1080  
Qy 775 CTGAGCTTCAATATGAAATCTTTCTGACATTTGATCAAGATGGAATCACTGACAGAGAA 834  
Db 1081 CTGGCTATATTTGAACTCGGCTGAGGTTGATGATGGAACAGCTTAAAGCAGAAAG 1140  
Qy 835 TTTATCTAGCTATGCACTTATTTGATGTTGCCATGTCTGTGACGACCTGCCGCCCTC 894  
Db 1141 TTTATCTTGGCAATGCACTTACTGACATGGCCAAAGCTGACAGCCTATTAACAAGTACT 1200  
Qy 895 CTGCTCTCAGAAATACATCCCTCCTTCTGAGAAAGTTGCTCCGAGTGGATGTC 954  
Db 1201 TTACCTCTGAGCTGTGTTCTCTCATCTTTCAAGAGGAAAGCAATTTGATTCATTAT 1260  
Qy 955 GTCATAGCTCTTTCTTGTGTGATTCAGAGGCTGCTGAGAGCCGTGTGAGAGATGAG 1014  
Db 1261 GGAACCTCGCTTATCATCAGAA-----AATGCAAGAA 1293  
Qy 1015 CACGAGCAGAGAAAGAACTGCTGACATTTGAAATTAAGAAAGCGGAGAACTTCGAG 1074  
Db 1294 GAGAGGCTTCAAGAAATTAACAGTTTACTTTGAGAGCAAAAGGAAAGCAACTATGAG 1353  
Qy 1075 CGAGGCACTGTGAGCTGAGAAAGCCGCGCAAGCCTCTTGGAGCAGCAGCGCAAGAG 1134  
Db 1354 CGAGGCACTGTGAGCTGAGAAAGCCGCGCAAGCCTCTTGGAGCAGCAGCGCAAGAG 1413  
Qy 1135 CAGAGCGGCTTGGCTGAGCTGAGAGCCGCGCAGAGAGGAAAGAGCGGCGAGCCGACG 1194  
Db 1414 GCAGAACTGTAAGCCGAGAAAGAAAGAGAGTGGAGCAAGAAACAGAGATTTACAA 1473  
Qy 1195 GAGCAGAGGCGCAAGCGGAGCTGAGCTGAGAAAGCAGCTGAGAGAGCAGCGGAGCTG 1254  
Db 1474 GAACTGAATGGAAGAAACAACTTGAATTAAGAAACCTTGAAGAGCAAGCGGATTTG 1533  
Qy 1255 GAGCGCAGGAGAGGAGGAGAGAGAGAGATCGAGAGCGCGAGCGCAAAACGG 1314  
Db 1534 GAGAGCAACGAGAGGAGAGAGAGAGAGATGAAAGAAAGAGAGAGCGAGCAAAACAG 1593  
Qy 1315 GAACTGGAAGGCGAGCAGCAACTTGAATGGGAACGGAACCGAGACAGAACTCTGAAT 1374  
Db 1594 GAACTTGAACGACAGCTCGCTTGAATGGAGAGAAATTTGGCGAGCAGAGCTTCCAA 1653  
Qy 1375 CAGAGAAAGAGAGAGAGAGAGCGGAGCCGTGCTCTGAGGCAAGAGAGAACTCTGAG 1434  
Db 1654 CAAAGAAATAGAGAAACAAAGAAATTTGTCAGTTAACTTAAAGAAAGAAATCTTCA 1713  
Qy 1435 TTTGAGTTAAGCTGAGATGACAAAGAAAGCATAGCTAAGAGAAATCTTCAGATATC 1494  
Db 1714 CTTGAGTTGAGAGCATGAAATGGCAACATCGAGATCTCAGAGAGACTTTCAGAGATG 1773  
Qy 1495 AGGTGTCAGCTGGCAACCGAGGCAAGAAATGAGAGCAAGAACATGTAGAGAGCTA 1554  
Db 1774 CGACTCAAAAGCAAACTCAAAAGACTGAGCTGGAAGTTCTGATTAAGCAGTGTGACTTG 1833  
Qy 1555 AGAATTTGCTGAATCAACCACTTACAGCAGAGTGGAGAAATCTCAGCAAAATGTTGGA 1614  
Db 1834 GAAATTTATGAAATCAAGCAACTTCAACAGAACTTCAAGAAATATCAGAAATGATTATC 1893  
Qy 1615 AGACTTTATTCAGAGAAAGATATCTCAGTACAGAGTAAAGCAAGCTCCGAGCAAGT 1674  
Db 1894 TATCTGCTACTGAGAGCAATTTAATGAAGAAATTAAGAAACATGCGCTTCACTAAC 1953

Qy 1675 TTGCTAGAGACTGCTCTTCTTACCTTCAAAAGAGCCTTGGAGCAAGAGCTGGCCCG 1734  
Db 1954 ACACCTGATTCAGGGGTCAAGTTTACTTATTAATAATCATTAAGAAAGAAATTAATG 2013  
Qy 1735 CAGCAGCTCCGGAGAGAGCTGAGAGAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1794  
Db 2014 CAAAGACTTAAAGAAAGAGTGAAGTCTTTGAAAGAAAGAGAGAGAGAGAGAGAGAG 2073  
Qy 1795 AATGATGTTTCAACAACAGCTGAA-----GGAAGTGAAGAGATACATATG 1842  
Db 2074 ATGATTTCTTTAATCATCACTAAGTGTGGAAATATGATGATGATCTGTTCTTCAAGTGC 2133  
Qy 1843 AAACGCAATCTCAGAAAGCAGAGGCTCCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1902  
Db 2134 CTTTGTCTGCTTGAAGCTGTCTCAACCACTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2193  
Qy 1903 GAGAGAAAGAGCTGAGATTAGAGAAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1956  
Db 2194 ACCTACAAACAGAGAGAGTATGAGCTTGAACAGCTTTATTAAGATCAAACTGACAAAGT 2253  
Qy 1957 CAGGAAAGAGCAAGCAATGCTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2016  
Db 2254 AAGGAAATTTGAAAGAGAAAGATTAGAACTAATGCAAGAAAGAAAGAGAGAGAGAG 2313  
Qy 2017 AAACCCACAG 2076  
Db 2314 GCAAG 2373  
Qy 2077 GAGAGAGCAAGCCGGAATATGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2136  
Db 2374 GAAAGAAACAAAG 2433  
Qy 2137 CCAGTAAAGCTGAGCAGCAG 2196  
Db 2434 CGGAAAGCTGAG 2493  
Qy 2197 TCTGACAG 2256  
Db 2494 CGTGAAG 2550  
Qy 2257 AGTCAAGTGAATCAACATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2316  
Db 2551 AACCATGATGAG 2610  
Qy 2317 GAG 2376  
Db 2611 GAG 2670  
Qy 2377 TATGAG 2436  
Db 2671 GAG 2721  
Qy 2437 ACATTCGCCCCCTGCCCAACCTGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2496  
Db 2722 AAGGCTTATCTTCTCTCAAGGTTTCTTATCTGTAACCTTCAACCTTCTTCAACCTT 2781  
Qy 2497 TCTTCTGAGCCCTCAGCAACCCCAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2556  
Db 2782 TCTTCAATCAACAGAGATCAGTGAATGATCAAAATGAT----- 2824  
Qy 2557 AGCTCAAG 2616  
Db 2825 -----CTTTTCA 2832  
Qy 2617 ACCGTAAGTATGCTGGCAGTTACAGAGAGATCAGCTTTTACCCAGCAGAGAGAGAGAG 2676  
Db 2833 AACCTAATCTGAATATCATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2892  
Qy 2677 GAGCTCTCCCATCTCCCTCTTGGCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2736  
Db 2893 GAG---ATCTGATACCTATTCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2949

Qy	2737	GCCTGATCCCTGGAGAGCCAAAGAAAGCAACACCTTAAATTTTAAACAAAGTACGCT	2796
Db	2250	GCCCTTGTCTCTGACGTGCAAGAAAGATTAACCACTTGAACCTTCTCAAAACATGACATT	3009
Qy	2797	ATCACCGCTTCGAAACAGCAAGACATGCTGCTGTTGGAGAGTTCAAGTCAGAAAGGT	2856
Db	3010	ATTACTGCTTGGAGAGCAAGAAATTTGTGGTTGGGAGAGTGCATGAGGAAGAGA	3069
Qy	2857	TGCTTCCCCAAGCTTACGTGAACCTCATTTTCAAGGCCCCGTAAAGAAATCCACAAGATC	2916
Db	3070	TGGTTTCCCAATCTTATGTCAAGATCATTTCTGGAGTGAAGTAAACGGGAA-----	3123
Qy	2917	GATACCTGGCCCTACTGAAAGTCCCTGCTAGCTTAAAGAGAGTGGCTTCCCGGCGCCAA	2976
Db	3124	-----GAAACCAAGACCTTGTATGCAAGCTGTAAATTAAGAAACCTACCTCGGACCTCAT	3177
Qy	2977	CCAGCCATTCCTCCGAGAGAGATTATTTGCAATGACATACATACAGAGATTCTGACAGAA	3036
Db	3178	TCAG-----TTGGAGAGAAATATATTGCACTTATTCATATTTCAAGTGTGGAACTTGA	3211
Qy	3037	GATTTAACCTTTACGCAAGGGGATGTGATTGTGTTAACCAAGAAAGATGTGACCTGTGG	3096
Db	3232	GATTTGACTTTCCACAGAAAGGTGAAGAAATATGTGTACCAAGAAAGATGAGAGTGTGG	3291
Qy	3097	ACGGGAACGGGCGGCGCAAGATCCGGAGTCTTCCCTTAACTATGTAGGCTTAAACAT	3156
Db	3292	ACAGGAAGTATTGGAGATTAAGTGAAGTATTTTCCATCAACTATATGTCAACCAAGAGAT	3351
Qy	3157	TCAGAGGCTCTGTGAACCTGCTGTGGAAAGCAAGGAGTTTAGAAGAAAGAACTGAATTTGCC	3216
Db	3352	CAAGAGAGTTTGTGGAGTGTCTACCAAGTCTGGAGCATCAAAATTAAGAAACCTGAGATTGCT	3411
Qy	3217	CAGGTTTATGCTTCTACGCTGCTACTGTGTCCTCCGAACAACCTACCTGTGCTTGGGAG	3276
Db	3412	CAGGTAACCTTCAGATATGTGTGCTTCTGTTCGAACAACTTAGCCTTGCAACAGAGCAG	3471
Qy	3277	CTGATTCGATCCGGAAGAAAGAACCCAGGAGATGCTGGGAAGAGAACTGCACACTCGA	3336
Db	3472	TTAATATTAATTTCTAAGAGAAATTAACAAGTGGGTGTGGCAAGAGATTAAGAGCCAGA	3531
Qy	3337	GGGAAAGCGCCAGATAGGAGTGTTCACAGCAAAATTAATGTCAAACTTTAAGCCCCGA	3396
Db	3532	GGAAAAAGGAGAAAGATGAGTGTTCCTGCGACATGTTAACTTTTGGGTCAAGT	3591
Qy	3397	ACAAAGCAAAATCACCCCACTGAGCTAACCAAGACCGCAGTGCAGCCAGCACTGTCCAG	3456
Db	3592	AGTGAAGAGCCACACCTGCTTTTCATC-----CTGTATGTGAG	3630
Qy	3457	GTGATCGGAGATGACGATTAACCGGCCGAAGACGATGACGAACTAGCTTAGCAAGAGGC	3516
Db	3631	GTGATTCCTATGTATGTACTGTACAGCAAAATTAAGAAAGATGAGCTCAGTTTCTCCAAAGGA	3690
Qy	3517	CAGATTCATCAACGTCCTCAACAAGAGAGACCCGCACTGTGTGAAAAGAGAGTCACTAGTGG	3576
Db	3631	CAACTCATTAATGTATGAACAAAGATGATCCTGATTTGTGTGCCAAGAGAGATTCACGGG	3750
Qy	3577	CAAGTGGGCTCTTCCATCAATTAATGTAAGCTGACCAAGACATGAGCCCAAGCCAG	3636
Db	3751	GTGACTGGTCTCTTCTCTCAAACTAGCTTAAGATGACAGACAGACTCAGATCCAAGTCAA	3810
Qy	3637	CAATG	3641
Db	3811	CAGTG	3815

```
Search completed: March 14, 2003, 08:17:28
Job time : 8944.04 secs
```



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2003, 01:28:45 ; Search time 650.685 Seconds  
(without alignments)  
12604.835 Million cell updates/sec

Title: US-09-674-237A-2  
Perfect score: 3642  
Sequence: 1 atggctcagttccacacc.....tggaccaccgacgaatga 3642

Scoring table: IDENTITY NUC  
Gapop 10\_0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues  
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: N.Geneseq.101002.\*  
2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
21: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3642	100.0	3723	21	Mouse Esei coding
2	3639.4	99.9	5144	21	Mouse Esei coding
3	3639.4	99.9	5738	21	Mouse Esei coding
4	3630	99.7	5082	21	Mouse Esei full le
5	2856	78.4	5195	20	Human SH3D1A cDNA
6	2831	77.7	5458	20	Human SH3D1A cDNA
7	2491.6	68.4	7435	23	DNA encoding novel
8	2429.6	66.7	5199	20	Human SH3D1A cDNA
9	2043.8	56.1	3319	23	DNA encoding novel

10	2043.8	56.1	3466	22	AA163825	Human polynucleoti
11	1495	41.0	2079	20	AA234573	Human SH3D1A cDNA
12	1483.8	40.7	2131	22	AAK94139	Human Full-length
13	1291.4	35.5	2874	23	AA584762	DNA encoding novel
14	1259.2	34.6	2131	22	AA165678	Human cDNA sequenc
15	1252.8	34.4	3231	20	AA234574	Human SH3D1A cDNA
16	1104.2	30.3	1676	22	AAK94611	Human SH3D1A cDNA
17	858	23.6	1389	17	AAK9795	Human SH3P17 gene.
18	721	19.8	5828	24	AA147247	Allergic disease e
19	713.8	19.6	6103	22	AAK52332	Human polynucleoti
20	677	18.6	4625	21	AA239010	Mouse Esei full le
21	677	18.6	4975	21	AA239027	Mouse Esei coding
22	677	18.6	6014	21	AA239026	Mouse Esei coding
23	666.6	18.3	3593	21	AA239011	Mouse Esei coding
24	648	17.8	4447	22	AA602055	DNA encoding molec
25	485.4	13.3	877	22	AAK91610	Human CDNA 5'-end
26	485.4	13.3	877	22	AAK93179	Human CDNA clone r
27	477.4	13.1	2017	21	AA69762	Human ovarian carc
28	477.4	13.1	2017	24	ABN72656	Human carcinoma
29	387.4	10.6	1329	22	AAH15280	Human CDNA sequenc
30	356	9.8	548	22	AA180000	Human polynucleoti
31	356	9.8	548	22	AA163919	Human polynucleoti
32	356	9.8	568	22	AA531621	DNA encoding nove
33	356	9.8	568	23	ABK43814	DNA encoding novel
34	340.8	9.4	2067	24	ABQ55007	Human ovarian anti
35	332.4	9.1	2873	17	AAK97959	Human clone 65 gen
36	331.4	9.1	3746	23	ABK43586	DNA encoding novel
37	331.4	9.1	4210	22	AA527090	DNA encoding nove
38	281	7.7	831	22	AAH03435	Human CDNA clone (
39	260	7.1	676	22	AAH08166	Human SH3P18 gene.
40	258	7.1	747	17	AAK97956	Human secreted exp
41	250	6.9	346	21	AA44038	DNA encoding novel
42	210.8	5.8	395	23	AA584759	Human breast cell
43	198.8	5.5	270	22	ABA46803	Human foetal liver
44	198.8	5.5	270	22	ABA46803	Probe #10273 for g
45	198.8	5.5	270	22	ABA11807	

## ALIGNMENTS

RESULT 1  
ID AA239009 standard; cDNA; 3723 BP.  
XX AA239009;  
AC  
XX  
XX 28-FEB-2000 (first entry)  
DT  
XX  
DE Mouse Esei coding sequence.  
XX  
XX Mouse; murine; Esei; Esei2; endocytosis; vesicular trafficking;  
KW regulation; actin cytoskeleton; detection; cancer; infection;  
KW EH-domain and SH-domain regulator of endocytosis; anticancer;  
KW antiproliferative; antiviral; ss.  
XX  
XX Mus sp.  
OS  
XX  
XX WO9955728-A2.  
PN  
XX  
XX 04-NOV-1999.  
PD  
XX  
XX 27-APR-1999; 99WO-CA00375.  
PF  
XX  
XX 27-APR-1998; 98CA-2230201.  
PR 05-FEB-1999; 99US-0118739.  
XX  
XX (HSCR-) HSC RES & DEV LP.  
PA  
XX  
XX Egan SE, Wang W, Sengar A;  
PI  
XX WPI, 2000-052802/04.  
DR  
XX P-PSDB; AAY57444.

XX New nucleic acid encoding Ese1 and 2 proteins, involved in regulation  
 PT of endocytosis, used e.g. for treating cancer or preventing viral  
 PT infection -

XX Claim 6; Page 40-42; 99pp; English.

XX The present sequence encodes mouse Ese1. The present invention  
 CC specifically describes mammalian Ese1 and 2 proteins (I) and their splice  
 CC variants (Ese = EH-domain and SH3-domain regulator of endocytosis). (I)  
 CC are involved in regulation of clathrin-mediated endocytosis (as a complex  
 CC with Eps15 protein), vesicular trafficking and actin cytoskeleton.  
 CC Generally (I) (or its (ant)agonists, mimetics, fragments and inactive  
 CC mutants); (I)-specific antibodies (Ab); sequences antisense to the (I)  
 CC polynucleotide; agents that downregulate expression of Ese genes or  
 CC antagonists of an Ese binding partner are used to treat diseases  
 CC associated with undesirable endocytosis and resulting changes in cellular  
 CC function. Particularly overexpression of Ese1 is used to block  
 CC clathrin-mediated endocytosis in vivo or in cell cultures, while  
 CC administration of (I) is used to promote endocytosis of selected cells.  
 CC (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of  
 CC cells that can be stimulated to proliferate by a growth factor receptor;  
 CC and similar compounds (also inactive Ese mutants) can be used to prevent  
 CC viral infection. Endocytosis may also be regulated, in vivo or in cell  
 CC cultures, by forming an Ese-Eps15 complex, then binding dynamn to the  
 CC complex. Generally conditions that can be treated include cancer;  
 CC abnormal cell division or migration; viral infection; or abnormal  
 CC receptor signalling, tissue development or synaptic transmission.  
 CC

SQ Sequence 3723 BP; 1065 A; 941 C; 1025 G; 692 T; 0 other;

Query Match 100.0%; Score 3642; DB 21; Length 3723;  
 Best Local Similarity 100.0%; Pred No. 0;  
 Matches 3642; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCTCAGTTTCCACACCTTTGGTGTAGCCTGTGATCTGTGGCCATTAACCTGTGAG 60  
 Db 1 ATGGCTCAGTTTCCACACCTTTGGTGTAGCCTGTGATCTGTGGCCATTAACCTGTGAG 60  
 QY 61 GAAAGGGCCCAAGCATGACCCAGCAGTCTTCTAGCTGAAAGCCGATAGGGGATTTTAACT 120  
 Db 61 GAAAGGGCCCAAGCATGACCCAGCAGTCTTCTAGCTGAAAGCCGATAGGGGATTTTAACT 120  
 QY 121 GGATGATCAAGGAGAGAACTTTTTCATCTGGGTTACTCAGCCTGTCTTAGCACA 180  
 Db 121 GGATGATCAAGGAGAGAACTTTTTCATCTGGGTTACTCAGCCTGTCTTAGCACA 180  
 QY 181 ATATGGGCGCTAGCGAATGATTAACGATGAGGAGATGATCAAGTGAATTTTCCATA 240  
 Db 181 ATATGGGCGCTAGCGAATGATTAACGATGAGGAGATGATCAAGTGAATTTTCCATA 240  
 QY 241 GCGATGAAGCTTTCAACTGAAGCTACAGATATAGTCCCTCCCACTTCCCTCCCT 300  
 Db 241 GCGATGAAGCTTTCAACTGAAGCTACAGATATAGTCCCTCCCACTTCCCTCCCT 300  
 QY 301 GTCATGAACAGCAACCGATGCTATTTCCAGTGCACAGCACTTTGTATAGAGGGATT 360  
 Db 301 GTCATGAACAGCAACCGATGCTATTTCCAGTGCACAGCACTTTGTATAGAGGGATT 360  
 QY 361 GCTAGCATGCGACCACTCAAGCTGTGCTCTGTGCGCAATGGCTTCATTCAGTTGTT 420  
 Db 361 GCTAGCATGCGACCACTCAAGCTGTGCTCTGTGCGCAATGGCTTCATTCAGTTGTT 420  
 QY 421 GGAATGCTCCACCTTATATCTTGTCTCTCCACAGACAGATGCTCCCTGGCTAAC 480  
 Db 421 GGAATGCTCCACCTTATATCTTGTCTCTCCACAGACAGATGCTCCCTGGCTAAC 480  
 QY 481 GGGGCTCTCCGTCATACAGCTCTGCTGCTTGGCATCTTGAGGCCACATGAGGCA 540  
 Db 481 GGGGCTCTCCGTCATACAGCTCTGCTGCTTGGCATCTTGAGGCCACATGAGGCA 540  
 QY 541 AAGAGTTCTTCTTCCAGCAGATCTGTCCAGGCTCAATTAACCTAAGTTACGAAAG 600  
 Db 541 AAGAGTTCTTCTTCCAGCAGATCTGTCCAGGCTCAATTAACCTAAGTTACGAAAG 600

Db 541 AAGAGTTCTTCTTCCAGCAGATCTGTCCAGGCTCAATTAACCTAAGTTACGAAAG 600  
 QY 601 GCACATCATTCATGATGTGCGAGGCCCTTCACAGCAGCAAGTGGCTGTGCTCACTCA 660  
 Db 601 GCACATCATTCATGATGTGCGAGGCCCTTCACAGCAGCAAGTGGCTGTGCTCACTCA 660  
 QY 661 TCAAGCTGAAATACAGCAGATTTATCAACAGCCACAGCAAAACTATAGTGAACACTTA 720  
 Db 661 TCAAGCTGAAATACAGCAGATTTATCAACAGCCACAGCAAAACTATAGTGAACACTTA 720  
 QY 721 ACAGTCCCGCAGGAGAGAACTATCTCATGCAATCAAGTTTACCAGGCTCAGCTGGCT 780  
 Db 721 ACAGTCCCGCAGGAGAGAACTATCTCATGCAATCAAGTTTACCAGGCTCAGCTGGCT 780  
 QY 781 TCAATATGGAATCTTTCTGACATGATGATGAGTGAATACTCACTGCAGAAAGATTATC 840  
 Db 781 TCAATATGGAATCTTTCTGACATGATGATGAGTGAATACTCACTGCAGAAAGATTATC 840  
 QY 841 CTAGCTATGACCTTAATTGATGTCATGTCGTCAGGCACTGCGCGCTGCTGCT 900  
 Db 841 CTAGCTATGACCTTAATTGATGTCATGTCGTCAGGCACTGCGCGCTGCTGCT 900  
 QY 901 CCAGATATCATCTCTCTCTTCTTCCAGAAAGATTGCTCCGCACTGGGATGTCCGTATA 960  
 Db 901 CCAGATATCATCTCTCTCTTCTTCCAGAAAGATTGCTCCGCACTGGGATGTCCGTATA 960  
 QY 961 AGCTTTCTTCTGTCGATCAAGGCTGCTGAGAGCCGTCGTAGAGATGAGCAGCAG 1020  
 Db 961 AGCTTTCTTCTGTCGATCAAGGCTGCTGAGAGCCGTCGTAGAGATGAGCAGCAG 1020  
 QY 1021 CCAGAGAGAAATCGCTGTCATTTGAAGATTAAGAGCGGAGAACTTCCAGCAGGCG 1080  
 Db 1021 CCAGAGAGAAATCGCTGTCATTTGAAGATTAAGAGCGGAGAACTTCCAGCAGGCG 1080  
 QY 1081 AGTGTGAGCTGAGAAAGCGCGCCAGCGCTTGTGAGAGAGCAGCCCAAGCAGAGAG 1140  
 Db 1081 AGTGTGAGCTGAGAAAGCGCGCCAGCGCTTGTGAGAGAGCAGCCCAAGCAGAGAG 1140  
 QY 1141 CGTTGGCTCAGCTGAGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
 Db 1141 CGTTGGCTCAGCTGAGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200  
 QY 1201 GAGGCGAAGCGGAGCTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
 Db 1201 GAGGCGAAGCGGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
 QY 1261 CAGCGAAG 1320  
 Db 1261 CAGCGAAG 1320  
 QY 1321 GAAAGGAG 1380  
 Db 1321 GAAAGGAG 1380  
 QY 1381 AAAG 1440  
 Db 1381 AAAG 1440  
 QY 1441 TTAAGAGCTGAGATGAGCAAAAGCATCAGCTAGAGAGAGAGAGAGAGAGAGAGAG 1500  
 Db 1441 TTAAGAGCTGAGATGAGCAAAAGCATCAGCTAGAGAGAGAGAGAGAGAGAGAGAG 1500  
 QY 1501 CGACTGAG 1560  
 Db 1501 CGACTGAG 1560  
 QY 1561 GCTGAATACCTTACAG 1620  
 Db 1561 GCTGAATACCTTACAG 1620  
 QY 1621 ATTCCAG 1680  
 Db 1621 ATTCCAG 1680





KW EH-domain and SH3-domain regulator of endocytosis; anticancer;  
 KW antiproliferative; antiviral; ss.  
 OS Mus sp.  
 XX WO955728-A2.  
 PN 04-NOV-1999.  
 PD 27-APR-1999; 99MO-CA00375.  
 PF 27-APR-1999; 98CA-2230201.  
 PR 05-FEB-1999; 99US-0118739.  
 XX (HSCR-) HSC RES & DEV LP.  
 PA Egan SE, Wang W, Sengar A;  
 PI WPI; 2000-052802/04.  
 DR P-PSDB; AAY57449.  
 XX  
 XX New nucleic acid encoding Esei and 2 proteins, involved in regulation  
 PT of endocytosis, used e.g. for treating cancer or preventing viral  
 PT infection -  
 PS Claim 6; Page 59-62; 99pp; English.  
 XX  
 XX The present invention specifically describes mammalian Esei and 2  
 CC proteins (I) and their splice variants (Ese = EH-domain and SH3-domain  
 CC regulator of endocytosis). (I) are involved in regulation of clathrin-  
 CC mediated endocytosis (as a complex with Eps15 protein), vesicular  
 CC trafficking and actin cytoskeleton. Generally (I) (or its (ant)agonists,  
 CC mimetics, fragments and inactive mutants); (I)-specific antibodies (Ab);  
 CC sequences antisense to the (I) polynucleotide; agents that downregulate  
 CC expression of Ese genes or antagonists of an Ese binding partner are  
 CC used to treat diseases associated with undesirable endocytosis and  
 CC resulting changes in cellular function. Particularly overexpression of  
 CC Esei is used to block clathrin-mediated endocytosis in vivo or in cell  
 CC cultures, while administration of (I) is used to promote endocytosis of  
 CC selected cells. (Ant)agonists of (I) or Ab are used to suppress abnormal  
 CC proliferation of cells that can be stimulated to proliferate by a growth  
 CC factor receptor; and similar compounds (also inactive Ese mutants) can be  
 CC used to prevent viral infection. Endocytosis may also be regulated, in  
 CC vivo or in cell cultures, by forming an Ese-Eps15 complex, then binding  
 CC dynam to the complex. Generally conditions that can be treated include  
 CC cancer; abnormal cell division or migration; viral infection; or abnormal  
 CC receptor signaling, tissue development or synaptic transmission. The  
 CC present sequence represents mouse Esei coding sequence.  
 XX  
 SO Sequence 5144 BP; 1447 A; 1298 C; 1415 G; 984 T; 0 other;  
 Query Match 99.9%; Score 3639.4; DB 21; Length 5144;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 3640; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 241 GCCATGAAGCTTATCAAACTGAAGCTAAGAGATATACGCTCCCTCCACACTTCCCT 300  
 QY 301 GTCATGAAGAGCAACAGTGGCTATTTCCAGTGCACAGCATTTTGTATAGAGGAT 360  
 Db 301 GTCATGAAGAGCAACAGTGGCTATTTCCAGTGCACAGCATTTTGTATAGAGGAT 360  
 QY 361 GCTAGCATGCCACACTCAAGCTGTTCCTCTGTGCAATGGGCTCCCAATTCAGTTT 420  
 Db 361 GCTAGCATGCCACACTCAAGCTGTTCCTCTGTGCAATGGGCTCCCAATTCAGTTT 420  
 QY 421 GGAATGTCTCCACCTAGTATCTTCTGTCCCTCCAGCAGCATGCTCCCTGCTTAAC 480  
 Db 421 GGAATGTCTCCACCTAGTATCTTCTGTCCCTCCAGCAGCATGCTCCCTGCTTAAC 480  
 QY 481 GGGCTCTCCCTGATATACAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
 Db 481 GGGCTCTCCCTGATATACAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
 QY 541 AAGATCTCTCTCCAGCAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600  
 Db 541 AAGATCTCTCTCCAGCAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600  
 QY 601 GCACATCATTCGATGTGCGCAGGCGCCCTCCAGCAGCATGAGGCTGCTCAGTCA 660  
 Db 601 GCACATCATTCGATGTGCGCAGGCGCCCTCCAGCAGCATGAGGCTGCTCAGTCA 660  
 QY 661 TCAGGCTGAATTCAGGCGCTTATTAACAGCCAGACAAACTATGATGACATTA 720  
 Db 661 TCAGGCTGAATTCAGGCGCTTATTAACAGCCAGACAAACTATGATGACATTA 720  
 QY 721 ACAGGCTCCAGGCGAAGATTCATGATGATGATGATGATGATGATGATGATGATGAT 780  
 Db 721 ACAGGCTCCAGGCGAAGATTCATGATGATGATGATGATGATGATGATGATGATGAT 780  
 QY 781 TCAATATGAATCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
 Db 781 TCAATATGAATCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
 QY 841 CTAGCTATGACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900  
 Db 841 CTAGCTATGACCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900  
 QY 901 CCAGATATCATCT 960  
 Db 901 CCAGATATCATCT 960  
 QY 961 AGCTCTTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020  
 Db 961 AGCTCTTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020  
 QY 1021 CCAAGAGAAAGAACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
 Db 1021 CCAAGAGAAAGAACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
 QY 1081 AGTGTGAGCTGAGAGAGAGCGCGCAAGCGCTCTTGAAGAGAGAGAGAGAGAGAG 1140  
 Db 1081 AGTGTGAGCTGAGAGAGAGCGCGCAAGCGCTCTTGAAGAGAGAGAGAGAGAGAG 1140  
 QY 1141 CGGTGTGCTGAGCTGAGAGAGAGCGCGCAAGCGCTCTTGAAGAGAGAGAGAGAGAG 1200  
 Db 1141 CGGTGTGCTGAGCTGAGAGAGAGCGCGCAAGCGCTCTTGAAGAGAGAGAGAGAGAG 1200  
 QY 1201 GAGGCTAAGCGGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
 Db 1201 GAGGCTAAGCGGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260  
 QY 1261 CAGCGAG 1320  
 Db 1261 CAGCGAG 1320  
 QY 1321 GAAAGGAG 1380  
 Db 1321 GAAAGGAG 1380

Qy	1381	AACAAGAGCAGAGGGCACCGTGTCTGAAAGCAAGAGGAAGACTCTGGAGTTTGG	1440
Db	1381	AACAAGAGCAGAGGGCACCGTGTCTGAAAGCAAGAGGAAGACTCTGGAGTTTGG	1440
Qy	1441	TTAGACCTCTGAATGACAAAAGACATCAGCTGGAAGGAAAACTTCAGGATTCAGTGT	1500
Db	1441	TTAGACCTCTGAATGACAAAAGACATCAGCTGGAAGGAAAACTTCAGGATTCAGTGT	1500
Qy	1501	CGACTGCGAACCCAGAGGCAAGAAATTGAGAGCCGAACAAGTCTTAGAGCTTAGAAT	1560
Db	1501	CGACTGCGAACCCAGAGGCAAGAAATTGAGAGCCGAACAAGTCTTAGAAGTT	1560
Qy	1561	GCTGAATTCACCACTTACAGCAGCAGTTCGAGGATTCACGAAATGTTGGAAGACTT	1620
Db	1561	GCTGAATTCACCACTTACAGCAGCAGTTCGAGGATTCACGAAATGCTTGGAGACTT	1620
Qy	1621	ATTCAGAGAAACAGATACCTCAGTGACAGATTAAAACAAGTCAGAGAAACGTTTGAT	1680
Db	1621	ATTCAGAGAAACAGATACCTCAGTGACAGATTAAAACAAGTCAGAGAAACGTTTGAT	1680
Qy	1681	AGAGACTCGCTTCTTACCCCTCAAAAGACCTTGGAAAGCAAGAGAGCTGGCCGACAG	1740
Db	1681	AGAGACTCGCTTCTTACCCCTCAAAAGACCTTGGAAAGCAAGAGAGCTGGCCGACAG	1740
Qy	1741	CTCCGGAGCAGCTGCGACGAGTGGAGAGAGACCAGTCAAAGTTCAGAGATTTAT	1800
Db	1741	CTCCGGAGCAGCTGCGACGAGTGGAGAGAGACCAGTCAAAGTTCAGAGATTTAT	1800
Qy	1801	GTTTTCACAACCAAGCTGGAAGAACTGAGAGGATTCATAGCAAAACAGCACTCCAGAG	1860
Db	1801	GTTTTCACAACCAAGCTGGAAGAACTGAGAGGATTCATAGCAAAACAGCACTCCAGAG	1860
Qy	1861	CAGAGTCCCTGAGGCGACGCACTGAAACAGAAAGACAGAGAGAGAGAGCCTTGAG	1920
Db	1861	CAGAGTCCCTGAGGCGACGCGCACTGAAAGCAAAAGACAGAGAGAGAGCCTTGAG	1920
Qy	1921	TTAGAGAAACAAAGGAAAGCGCTCAGAGCGAGTTCAGGAAAGGAGCAAGCATGCTG	1980
Db	1921	TTAGAGAAAGCAAAAGGAAAGCGCTCAGAGCGAGTTCAGGAAAGGAGCAAGCATGCTG	1980
Qy	1981	GAGCATGCGAGCAGAGGAGCGACGCCGCCGGAACCCACAGAGAGACAGACTG	2040
Db	1981	GAGCATGTGAGCAGAGGAGCGACGCCGCCGGAACCCACAGAGAGAGCAGACTG	2040
Qy	2041	AAGAGGAAGACAGTGTCCAGAGAAAGAGAGCGAGAGACAGACCAGCGCGAAATTCGA	2100
Db	2041	AAGAGGAAGACAGTGTCCAGAGAAAGAGAGCGAGAGAGAGAGACCAGCGCGAAATTCGA	2100
Qy	2101	GACAAAGCAGAGTGGCTTTTCCATCCGCACTCAGAGCCAGCTAAGCTGGCCAACCGACA	2160
Db	2101	GACAAAGCAGAGTGGCTTTTCCATCCGCACTCAGAGCCAGCTAAGCTGGCCAACCGACA	2160
Qy	2161	CCCTGTGTACACACAGAGAAAGCCCGCTTACATTTCTGCACAGAGAGGTAAAGTG	2220
Db	2161	CCCTGTGTACACACAGAGAAAGCCCGCTTACATTTCTGCACAGAGAGGTAAAGTG	2220
Qy	2221	GTAATTTAACGAGGCGCTGACCCCTTTGAAATCAGAAAGTCAAGATGAGATACCAATCG	2280
Db	2221	GTAATTTAACGAGGCGCTGACCCCTTTGAAATCAGAAAGTCAAGATGAGATACCAATCG	2280
Qy	2281	CCAGAGATATAGTCATGCTGATGAAAGCCAGACTGAGAGCCAGAGAGGCTTGGAGGA	2340
Db	2281	CCAGAGATATAGTCATGCTGATGAAAGCCAGACTGAGAGCCAGAGAGGCTTGGAGGA	2340
Qy	2341	GAGCTGAAAGGAGAGCGGATGGTTCCTGCAAACTATGCAAGAAAGATTCCAGAAAT	2400
Db	2341	GAGCTGAAAGGAGAGCGGATGGTTCCTGCAAACTATGCAAGAAAGATTCCAGAAAT	2400
Qy	2401	GAGGTCCTCACTCCAGCCAAACAGATGACCTGACATCTGCCCCCAACCTG	2460
Db	2401	GAGGTCCTCACTCCAGCCAAACAGATGACCTGACATCTGCCCCCAACCTG	2460

Oy	2461	GCCTGGGTAGAACCCCTGCTCTTTGGCCAGAGACCTCTTGTAGACCCTCCACAACCCC	2520
Db	2461	GCTTGCGTGAAGACCCCTGCTCTTTGGCCAGAGACCTCTTGTAGACCCTCCACAACCCC	2520
Oy	2521	AACAACCTGGGCAACACTTCAAGTTTCAAGTGGCCACAGCTCAAAAGAGACCGAAGACG	2580
Db	2521	AACAACCTGGGCAACACTTCAAGTTTCAAGTGGCCACAGCTCAAAAGAGACCGAAGACG	2580
Oy	2581	GACAACCTGGGATACGTGGGGGGCTCAGCCTTCTTGACCGTACCTAGTGTGGCCAGTTA	2640
Db	2581	GACAACCTGGGATACGTGGGGGGCTCAGCCTTCTTGACCGTACCTAGTGTGGCCAGTTA	2640
Oy	2641	CGGAGAGATACACCTTTTACCACGACAGGCACTGCTCTCCCAATCTCCGCTCTG	2700
Db	2641	CGGAGAGATACACCTTTTACCACGACAGGCACTGCTCTCTCCCAATCTCCGCTCTG	2700
Oy	2701	GGCCAGGAGTAAAGGTGGAAGGGCTACAGGCGCAAGCCCTGTATCCCTGGAGAGCCAAA	2760
Db	2701	GGCCAGGAGTAAAGGTGGAAGGGCTACAGGCGCAAGCCCTGTATCCCTGGAGAGCCAAA	2760
Oy	2761	AAAGACACCACTTAAATTTTAACAAAAGTGACGTATCACCGTCTTGGAACAGCAAGAC	2820
Db	2761	AAAGACACCACTTAAATTTTAACAAAAGTGACGTATCACCGTCTTGGAACAGCAAGAC	2820
Oy	2821	ATGTGTGTGTTTGGAGAAAGTTCAAGGTCAAGAGGTGTGTTCCCAACTCTTAAGCTAAA	2880
Db	2821	ATGTGTGTGTTTGGAGAAAGTTCAAGGTCAAGAGGTGTGTTCCCAACTCTTAAGCTAAA	2880
Oy	2881	CTCATTTTCAGGGCCCGTAAAGAAATCAACAACATCTGATCTGGCCCTACTAGAAAGTCT	2940
Db	2881	CTCATTTTCAGGGCCCGTAAAGAAATCAACAACATCTGATCTGGCCCTACTAGAAAGTCT	2940
Oy	2941	GCTAGTCTAAAGAGAGTGGCTTCCCGGCGCCAAAGCCACCATTCGGGAGAAAGATT	3000
Db	2941	GCTAGTCTAAAGAGAGTGGCTTCCCGGCGCCAAAGCCACCATTCGGGAGAAAGATT	3000
Oy	3001	ATTGCCATGTACACATACAGAAATTTGTGAGCAAGAGATTTAACTTTACACAAGGGAT	3060
Db	3001	ATTGCCATGTACACATACAGAAATTTGTGAGCAAGAGATTTAACTTTACACAAGGGAT	3060
Oy	3061	GTGATTTGTGTTACCAAGAAAGTGTGACTGTGACGGGAACGCTGTGGCGACAAAGTCC	3120
Db	3061	GTGATTTGTGTTACCAAGAAAGTGTGACTGTGACGGGAACGCTGTGGCGACAAAGTCC	3120
Oy	3121	GGAGTCTTCCCTTCTAACTATGTGAGGCTTAAAGATTCAAGAGGCTCTGGAACCTGTGG	3180
Db	3121	GGAGTCTTCCCTTCTAACTATGTGAGGCTTAAAGATTCAAGAGGCTCTGGAACCTGTGG	3180
Oy	3181	AAAAACAGGAGCTTGAAGAAAAAACTGAAATTTGCCACGATTATTCCTTCTAGCGTCT	3240
Db	3181	AAAAACAGGAGCTTGAAGAAAAAACTGAAATTTGCCACGATTATTCCTTCTAGCGTCT	3240
Oy	3241	ACTGTGTCGGAACAATCACACCTGGGCTCTGTGGACAGCTGATTCGATATCCGAAAAAGAC	3300
Db	3241	ACTGTGTCGGAACAATCACACCTGGGCTCTGTGGACAGCTGATTCGATATCCGAAAAAGAC	3300
Oy	3301	CCAGGTGATGTGTGGGAAGAGAACTGCAAGCTCGAGGAAAAAAGCGCAGATAGGTGG	3360
Db	3301	CCAGGTGATGTGTGGGAAGAGAACTGCAAGCTCGAGGAAAAAAGCGCAGATAGGTGG	3360
Oy	3361	TTTTCACCAAAATTTATGTCAAACTTTCAACCCCGGAACAAGCAAAATCACCCCAACTAG	3420
Db	3361	TTTTCACCAAAATTTATGTCAAACTTTCAACCCCGGAACAAGCAAAATCACCCCAACTAG	3420
Oy	3421	CTACCCAAAGACCGAGTGCAGCCAGAGTGTCCACAGTGTATGGGATATACGATTTACCC	3480
Db	3421	CTACCCAAAGACCGAGTGCAGCCAGAGTGTCCACAGTGTATGGGATATACGATTTACCC	3480
Oy	3481	GCCCGAAGCATGACGAATCTAGCTTTCACCAAAAGGCGAGATCATTAAGTCTCTCAACAG	3540
Db	3481	GCCCGAAGCATGACGAATCTAGCTTTCACCAAAAGGCGAGATCATTAAGTCTCTCAACAG	3540
Oy	3541	GAGGACCCGGAATGTGTGGAAGAGAAATGATGTGGCAAGTTTGGCTCTTCCCATCCAAAT	3600

Db 3541 GAGGACCCGACCTGTGGAAAGAGAGTCACTGGGCAATTTGGCTTCCCATCCAAAT 3600  
Qy 3601 TAGTAAAGCTGACCAAGACATGAGACCCAGCCAGCAATG 3641  
Db 3601 TAGTAAAGCTGACCAAGACATGAGACCCAGCCAGCAATG 3641

RESULT 3

AAZ39024  
ID AAZ39024 standard; cDNA; 5738 BP.

AC AAZ39024;

DT 28-FEB-2000 (first entry)

XX Mouse Esell cDNA sequence.

XX Mouse; murine; Esel; Esee; endocytosis; vesicular trafficking;  
KW regulation; actin cytoskeleton; detection; cancer; infection;  
KW EH-domain and SH3-domain regulator of endocytosis; anticancer;  
KW antiproliferative; antiviral; ss.

OS Mus sp.

PN MO9955728-A2.

PD 04-NOV-1999.

PF 27-APR-1999; 99MO-CA00375.

PR 27-APR-1998; 98CA-2230201.

PR 05-FEB-1999; 99US-0118739.

XX (HSCR-) HSC RES & DEV LP.

PI Egan SE, Wang W, Sengar A;

DR WPI; 2000-052802/04.

XX P-PSDB; AAY57449.

PT New nucleic acid encoding Esel and 2 proteins, involved in regulation  
of endocytosis, used e.g. for treating cancer or preventing viral  
infection -

XX Claim 6; Page 56-59; 99pp; English.

PS The present invention specifically describes mammalian Esel and 2  
CC proteins (I) and their splice variants (Ese = EH-domain and SH3-domain  
CC regulator of endocytosis). (I) are involved in regulation of clathrin-  
CC mediated endocytosis (as a complex with Eps15 protein), vesicular  
CC trafficking and actin cytoskeleton. Generally (I) (or its (ant)agonists,  
CC mimetics, fragments and inactive mutants); (I)-specific antibodies (Ab);  
CC sequences antisense to the (I) polynucleotide; agents that downregulate  
CC expression of Ese genes or antagonists of an Ese binding partner are  
CC used to treat diseases associated with undesirable endocytosis and  
CC resulting changes in cellular function. Particularly, overexpression of  
CC Esel is used to block clathrin-mediated endocytosis in vivo or in cell  
CC cultures, while administration of (I) is used to promote endocytosis of  
CC selected cells. (Ant)agonists of (I) or Ab are used to suppress abnormal  
CC proliferation of cells that can be stimulated to proliferate by a growth  
CC factor receptor; and similar compounds (also inactive Ese mutants) can be  
CC used to prevent viral infection. Endocytosis may also be regulated, in  
CC vivo or in cell cultures, by forming an Ese-Eps15 complex, then binding  
CC dynamin to the complex. Generally conditions that can be treated include  
CC cancer; abnormal cell division or migration; viral infection; or abnormal  
CC receptor signaling, tissue development or synaptic transmission. The  
CC present sequence represents mouse Esell cDNA sequence.

SO Sequence 5738 BP; 1556 A; 1456 C; 1613 G; 1113 T; 0 other;

Query Match 99.9%; Score 3639.4; DB 21; Length 5738;  
Best Local Similarity 100.0%; Pred. No. 0;

Matches 3640; Conservative 0; Mismatches 1; Indels 0; Gaps 0;			
Qy	1	ATGGCTAGTTTCCCAACACCTTTGGTGTAGCTGTGATGTCGAGGCGCATTAAGTGTGAG	60
Db	260	ATGGCTAGTTTCCCAACACCTTTGGTGTAGCTGTGATGTCGAGGCGCATTAAGTGTGAG	319
Qy	61	GAAAGGCGCAAGCATGACAGCACTTCTTAAGCTTGAAGCGGATAGGGGATTTATTACT	120
Db	320	GAAAGGCGCAAGCATGACAGCACTTCTTAAGCTTGAAGCGGATAGGGGATTTATTACT	379
Qy	121	GGTATGCAAGGAGGAGACTTTTTCCTCAATCTGGGTACCTGCTGTTAGCAAA	180
Db	380	GGTATGCAAGGAGGAGACTTTTTCCTCAATCTGGGTACCTGCTGTTAGCAAA	439
Qy	181	ATATGGGCGCTAGCGGACATGATATAGATGGAAGATGATCAAGTGAATTTTCCATA	240
Db	440	ATATGGGCGCTAGCGGACATGATATAGATGGAAGATGATCAAGTGAATTTTCCATA	499
Qy	241	GCCATGAAGCTTTATCAAACTGAAGCTACAGGATATAGCTTCCCTCCACATTTCCCT	300
Db	500	GCCATGAAGCTTTATCAAACTGAAGCTACAGGATATAGCTTCCCTCCACATTTCCCT	559
Qy	301	GTCATGAAGCAAGCAACGAGCTATTTCAGTGCACAGCATTTGGTATAGAGGAT	360
Db	560	GTCATGAAGCAAGCAACGAGCTATTTCAGTGCACAGCATTTGGTATAGAGGAT	619
Qy	361	GCTAGCATGCCACCACTGACAGCTGTGCTCTGTGCGAATGGGCTCCATTCAGTTGTT	420
Db	620	GCTAGCATGCCACCACTGACAGCTGTGCTCTGTGCGAATGGGCTCCATTCAGTTGTT	679
Qy	421	GGAATGTCGACCCCTTATGATCTTCTGCTCCCTCAGAGAGAGGCTCCCTGGCTAAC	480
Db	680	GGAATGTCGACCCCTTATGATCTTCTGCTCCCTCAGAGAGAGGCTCCCTGGCTAAC	739
Qy	481	GGGGCTCTCCCTGATCATACAGCTCTGCTGCTGCTTGGGATCCTGACGACATGCGCA	540
Db	740	GGGGCTCTCCCTGATCATACAGCTCTGCTGCTGCTTGGGATCCTGACGACATGCGCA	799
Qy	541	AAGAGTTCTTCTTCTGACAGATCTGCTCAGAGGTCACATTTAACTAATTACAGAG	600
Db	800	AAGAGTTCTTCTTCTGACAGATCTGCTCAGAGGTCACATTTAACTAATTACAGAG	859
Qy	601	GCACATCATTCGATGATGTCGAGGCGCCCTCCAGCAGAGATGGGCTGTCCTCACTCA	660
Db	860	GCACATCATTCGATGATGTCGAGGCGCCCTCCAGCAGAGATGGGCTGTCCTCACTCA	919
Qy	661	TCAAGGCTGAATATACAGGCACTTATTCACACGACGACAAACTATGATGACACTTA	720
Db	920	TCAAGGCTGAATATACAGGCACTTATTCACACGACGACAAACTATGATGACACTTA	979
Qy	721	ACAGGTCGCCAGGCAAGAACTATTCATGCAATCAAGTTAACCCAGGCTCAGCTGGCT	780
Db	980	ACAGGTCGCCAGGCAAGAACTATTCATGCAATCAAGTTAACCCAGGCTCAGCTGGCT	1039
Qy	781	TCAATATGAATCTTCTGACATGATCAAGATGAAAACTCAGTCGAGAGAAATTTATC	840
Db	1040	TCAATATGAATCTTCTGACATGATCAAGATGAAAACTCAGTCGAGAGAAATTTATC	1099
Qy	841	CTAGCTATGACCTTAATGATGTTGCCATGTCGTGTGTCAGCCATCCGCTCTGCT	900
Db	1100	CTAGCTATGACCTTAATGATGTTGCCATGTCGTGTGTCAGCCATCCGCTCTGCT	1159
Qy	901	CCAGAAATCATCCCTCTTCTTCCAGAGAGTTCGCTCCGCGAGTGGATTCCTGATATA	960
Db	1160	CCAGAAATCATCCCTCTTCTTCCAGAGAGTTCGCTCCGCGAGTGGATTCCTGATATA	1219
Qy	961	AGCTCTTCTTCTGATGATCAAGAGCTGCTGAGAGCGCTGCTCAAGAGATGAGCAGAG	1020
Db	1220	AGCTCTTCTTCTGATGATCAAGAGCTGCTGAGAGCGCTGCTCAAGAGATGAGCAGAG	1279
Qy	1021	CCAGAGAAAGAACTCCCTGCTGATCAATTAAGATTAAGAGCGGGAACCTTGAGCGAGGC	1080
Db	1280	CCAGAGAAAGAACTCCCTGCTGATCAATTAAGATTAAGAGCGGGAACCTTGAGCGAGGC	1339

QY 1081 AGTGTGAGCTGAGAAAGCGCGCAAGCGCTTTGAGACAGCGGCAAGAAGCAGAG 1140  
DB 1340 AGTGTGAGCTGAGAAAGCGCGCAAGCGCTTTGAGACAGCGGCAAGAAGCAGAG 1399  
QY 1141 CGGTTGGCTCAGCTGAGACGCGCGCGAGCAGAGAAAGAAAGCGGGAAGCGCCAGAGCAG 1200  
DB 1400 CGGTTGGCTCAGCTGAGACGCGCGCGAGCAGAGAAAGAAAGCGGGAAGCGCCAGAGCAG 1459  
QY 1201 GAGGCCAAGCGGAGCTGAGAGTGAAGACAGCTGAGAACGAGCGGAGCTGAGACCG 1260  
DB 1460 GAGGCCAAGCGGAGCTGAGAGTGAAGACAGCTGAGAACGAGCGGAGCTGAGACCG 1519  
QY 1261 CAGCGAGAGGAGAGAGAGAAAGAGATCGAGAGCGCGAGCGCGCAAAACCGGAATCTG 1320  
DB 1520 CAGCGAGAGGAGAGAGAGAAAGAGATCGAGAGCGCGAGCGCGCAAAACCGGAATCTG 1579  
QY 1321 GAAAGGACGAGAACTTGAATGGGAACGGAACCGAGACAGGAATCTTGAATCGAGAG 1380  
DB 1580 GAAAGGACGAGAACTTGAATGGGAACGGAACCGAGACAGGAATCTTGAATCGAGAG 1639  
QY 1381 AACAAAGAGAGAGAGAGCGGACCGTGTCTGAAAGCAAGAGAGAACTCTGAGTTGAG 1440  
DB 1640 AACAAAGAGAGAGAGAGCGGACCGTGTCTGAAAGCAAGAGAGAACTCTGAGTTGAG 1699  
QY 1441 TTGAAGCTCTGAATGACAAAAGACATCAGCTAGAAAGAAAATTTCAGGATATCAGGTGT 1500  
DB 1700 TTGAAGCTCTGAATGACAAAAGACATCAGCTAGAAAGAAAATTTCAGGATATCAGGTGT 1759  
QY 1501 CGACTGGCAACCCAGAGGCAAGAAATTGAGAGACAGAACAGTCTTAGAGAGCTAAGAAATT 1560  
DB 1760 CGACTGGCAACCCAGAGGCAAGAAATTGAGAGACAGAACAGTCTTAGAGAGCTAAGAAATT 1819  
QY 1561 GCGAAGATCAACCCACTTACAGCAGAGTTGAGAGAAATCTCAGCAATGCTTGAAGACTT 1620  
DB 1820 GCGAAGATCAACCCACTTACAGCAGAGTTGAGAGAAATCTCAGCAATGCTTGAAGACTT 1879  
QY 1621 ATTCCAGAGAAACAGATACTCAGTGAACAGTTTAAAACAAGTCAGCAGAACAGTTTGCAT 1680  
DB 1880 ATTCCAGAGAAACAGATACTCAGTGAACAGTTTAAAACAAGTCAGCAGAACAGTTTGCAT 1939  
QY 1681 AGAGATCTCGCTTTTACCTCTCAAAAAGCCTTGAAGCAAAAGAGCTGGCCGCGAGCAG 1740  
DB 1940 AGAGATCTCGCTTTTACCTCTCAAAAAGCCTTGAAGCAAAAGAGCTGGCCGCGAGCAG 1999  
QY 1741 CTCGGGAGAGCTGAGACGAGTGAAGAGAGAGACAGAGTCAAGCTGAGAGAGATTGAT 1800  
DB 2000 CTCGGGAGAGCTGAGACGAGTGAAGAGAGAGACAGAGTCAAGCTGAGAGAGATTGAT 2059  
QY 1801 GTTTTCAACCAACAGCTGAAGAACTGAGAGATACATAGCAACAGCACTCCAGAAAG 1860  
DB 2060 GTTTTCAACCAACAGCTGAAGAACTGAGAGATACATAGCAACAGCACTCCAGAAAG 2119  
QY 1861 CAGAGGCTCTGAGAGCAGCGGCACTGAAAGCAGAAAGACAGAGAGAGAAAGCTTGAG 1920  
DB 2120 CAGAGGCTCTGAGAGCAGCGGCACTGAAAGCAGAAAGACAGAGAGAGAAAGCTTGAG 2179  
QY 1921 TTGAGAGAGCAAAAGAGAGCGCTCAGAGACGAGTTCAAGAAAAGGAGCAAGCAATGGCTG 1980  
DB 2180 TTGAGAGAGCAAAAGAGAGCGCTCAGAGACGAGTTCAAGAAAAGGAGCAAGCAATGGCTG 2239  
QY 1981 GAGCATGTGAGCAGAGAGAGCAGCCAGCCCGGAAAACCCCAAGAGAGAGAGAGAGCTG 2040  
DB 2240 GAGCATGTGAGCAGAGAGAGCAGCCAGCCCGGAAAACCCCAAGAGAGAGAGAGAGCTG 2299  
QY 2041 AAGAGGAAAGACAGTGTCAAGAAAGAGAGCGGAAAGAGAGCCAAAGCCGGAATGCAA 2100  
DB 2300 AAGAGGAAAGACAGTGTCAAGAAAGAGAGCGGAAAGAGAGCCAAAGCCGGAATGCAA 2359  
QY 2101 GACAGAGAGTGGGCTTTTCCATCCGCAATCAGAGCCAGCTAAGTGGCCAGCCAGGCA 2160  
DB 2360 GACAGAGAGTGGGCTTTTCCATCCGCAATCAGAGCCAGCTAAGTGGCCAGCCAGGCA 2419

QY 2161 CCTGTCTACACAGAGAAAGCGCGCTTACATTTCTGACAGAGAGTGAAGATG 2220  
DB 2420 CCTGTCTACACAGAGAAAGCGCGCTTACATTTCTGACAGAGAGTGAAGATG 2479  
QY 2221 GATATTTACGAGGCGGTGATCCCTTTGAATCCGAAGTACCGATGAGATCACATCCAG 2280  
DB 2480 GATATTTACGAGGCGGTGATCCCTTTGAATCCGAAGTACCGATGAGATCACATCCAG 2539  
QY 2281 CCAGAGATATATGTCATGTTGATGAAGAGCAGATGAGAGGCGAGATGGCTTGAAGA 2340  
DB 2540 CCAGAGATATATGTCATGTTGATGAAGAGCAGATGAGAGGCGAGATGGCTTGAAGA 2599  
QY 2341 GAGCTGAAAGGGAAGACGGAGTGGTCCCTGCAAACTATGACAGAAAGATTCCAGAAAT 2400  
DB 2600 GAGCTGAAAGGGAAGACGGAGTGGTCCCTGCAAACTATGACAGAAAGATTCCAGAAAT 2659  
QY 2401 GAGGTTCCACTCCAGCCAAACAGTGAACCGATCTGACATCTGCCCCCTGCCCAACTG 2460  
DB 2660 GAGGTTCCACTCCAGCCAAACAGTGAACCGATCTGACATCTGCCCCCTGCCCAACTG 2719  
QY 2461 GCTCTGCGTGAAGACCCCTGCTCTTTGCAAGTGAACCTTCTGAGCCCTCCACAAACCCC 2520  
DB 2720 GCTCTGCGTGAAGACCCCTGCTCTTTGCAAGTGAACCTTCTGAGCCCTCCACAAACCCC 2779  
QY 2521 AACAACTGGGAGACTTCACTTCAGTTCAGGTGGCCAGAGCTCAACGAGAAACG 2580  
DB 2780 AACAACTGGGAGACTTCACTTCAGTTCAGGTGGCCAGAGCTCAACGAGAAACG 2839  
QY 2581 GACAACTGGGATATAGTGGGCGGCTCAGCTTCTTGACCGTACTAATGCTGGCCAGTTA 2640  
DB 2840 GACAACTGGGATATAGTGGGCGGCTCAGCTTCTTGACCGTACTAATGCTGGCCAGTTA 2899  
QY 2641 CGGAGAGATCAGGCTTTTACCCGAGCAGCCAGCCAGCTGCTCCCTCCGATCTCCGCTCTG 2700  
DB 2900 CGGAGAGATCAGGCTTTTACCCGAGCAGCCAGCCAGCTGCTCCCTCCGATCTCCGCTCTG 2959  
QY 2701 GCGCAGGCTGAAAAGTGAAGAGGCTTCAAGCGCAAGCCCTGTATTCCTTGAAGCCAAA 2760  
DB 2960 GCGCAGGCTGAAAAGTGAAGAGGCTTCAAGCGCAAGCCCTGTATTCCTTGAAGCCAAA 3019  
QY 2761 AAAGACAAACCACTTAAATTTTAAACAAAAGTACGTATCACCGTTCTGGAACAGCAAGAC 2820  
DB 3020 AAAGACAAACCACTTAAATTTTAAACAAAAGTACGTATCACCGTTCTGGAACAGCAAGAC 3079  
QY 2821 ATGTGTGTGTTGAGAAAGTTCAAGTCAAGAGGTTGTTCCCAAGTCTTACGTGAAA 2880  
DB 3080 ATGTGTGTGTTGAGAAAGTTCAAGTCAAGAGGTTGTTCCCAAGTCTTACGTGAAA 3139  
QY 2881 CTCAATTTCAAGGCGCGTGAAGAAATCCACAAGCATCGATCTGAGCCCTACTGAAAGTCT 2940  
DB 3140 CTCAATTTCAAGGCGCGTGAAGAAATCCACAAGCATCGATCTGAGCCCTACTGAAAGTCT 3199  
QY 2941 GCTAGTCTTAAAGAGATGCTTCCCGCGCGCCCAAGCAGCATTTCCCGAGAAAGATT 3000  
DB 3200 GCTAGTCTTAAAGAGATGCTTCCCGCGCGCCCAAGCAGCATTTCCCGAGAAAGATT 3259  
QY 3001 ATTGCAATGTACATACGAGAGTTCTGAGACAGAGATTTAACTTTACAGCAAGGGGAT 3060  
DB 3260 ATTGCAATGTACATACGAGAGTTCTGAGACAGAGATTTAACTTTACAGCAAGGGGAT 3319  
QY 3061 GTGATTGTGTTTACCAAGAAAGATGTGACTGTGAGCGGAAACGATGGGAGCAAGTCC 3120  
DB 3320 GTGATTGTGTTTACCAAGAAAGATGTGACTGTGAGCGGAAACGATGGGAGCAAGTCC 3379  
QY 3121 GGAAGTCTTCCCTTTAACTATGTGAGGCTTAAAGATTCAAGAGGCTTTGAACTGCTGGG 3180  
DB 3380 GGAAGTCTTCCCTTTAACTATGTGAGGCTTAAAGATTCAAGAGGCTTTGAACTGCTGGG 3439  
QY 3181 AAAACAGGAGTTTAAAGAAAAAACTGAAATTGCGCAGGTTATGCTTCTCGCTGCT 3240  
DB 3440 AAAACAGGAGTTTAAAGAAAAAACTGAAATTGCGCAGGTTATGCTTCTCGCTGCT 3499  
QY 3241 ACTGGTCCGAACAACCTACCCCTGGCTCTGGGCAAGCTGATCTGATCCGAAAAAGAAC 3300

Db 3500 ACTGATCCGAACTACCTCTGGCTGCTGGAGTGTCTGATCCGAAAGAAAC 3559  
 QY 3301 CCAAGTGTATGTGGGAAAGAACTGCAAGCTCCAGGAAAAAGCCAGATAGGTTG 3360  
 Db 3560 CCAAGTGTATGTGGGAAAGAACTGCAAGCTCCAGGAAAAAGCCAGATAGGTTG 3619  
 QY 3361 TTTCAGCAATATGTCAAACTTCTAAGCCCCGAAACAAGAAATCAACCCCACTGAG 3420  
 Db 3620 TTTCAGCAATATGTCAAACTTCTAAGCCCCGAAACAAGAAATCAACCCCACTGAG 3679  
 QY 3421 CTACCCAGAACCCGAGTGCAGCCAGCAGTGTCCAGTGTATGCGATGTCACACC 3480  
 Db 3680 CTACCCAGAACCCGAGTGCAGCCAGCAGTGTGTCCAGTGTATGTCACACC 3739  
 QY 3481 GCCCAGAACGATGACCACTAGCTTCTGAGCAAGGCGAGATCATCAACCTCCACAG 3540  
 Db 3740 GCCCAGAACGATGACCACTAGCTTCTGAGCAAGGCGAGATCATCAACCTCCACAG 3799  
 QY 3541 GAGGACCCGAGTGTGGAAGAGAGAGTCAAGTGGCAAGTGGCTCTTCCCATCAAT 3600  
 Db 3800 GAGGACCCGAGTGTGGAAGAGAGAGTCAAGTGGCAAGTGGCTCTTCCCATCAAT 3859  
 QY 3601 TATGTAAAGCTGACCAAGACATGAGACCCAGCCAGCAATG 3641  
 Db 3860 TATGTAAAGCTGACCAAGACATGAGACCCAGCCAGCAATG 3900

## RESULT 4

AAZ39008 standard; cDNA; 5082 BP.

AAZ39008;

28-FEB-2000 (first entry)

Mouse Ese1 full length cDNA sequence.

XX Mouse; murine; Ese1; Ese2; endocytosis; vesicular trafficking;  
 KW regulation; actin cytoskeleton; detection; cancer; infection;  
 KM EH-domain and SH3-domain regulator of endocytosis; anticancer;  
 KM antiproliferative; antiviral; ss.

Mus sp.

W0955728-A2.

04-NOV-1999.

27-APR-1999; 99MO-CA00375.

27-APR-1998; 98CA-2230201.

05-FEB-1999; 99US-0118739.

(HSCR-) HSC RES &amp; DEV LP.

Egan SE, Wang W, Sengar A;

WPI; 2000-052802/04.

P-PSDB; AAY57444.

New nucleic acid encoding Ese1 and 2 proteins, involved in regulation

infection -

Claim 6; Page 38-40; 99pp; English.

CC The present sequence encodes mouse Ese1. The present invention  
 CC specifically describes mammalian Ese1 and 2 proteins (1) and their splice  
 CC variants (Ese = EH-domain and SH3-domain regulator of endocytosis). (1)  
 CC are involved in regulation of clathrin-mediated endocytosis (as a complex  
 CC with Eps15 protein), vesicular trafficking and actin cytoskeleton.  
 CC Generally (1) (or its (ant)agonists, mimetics, fragments and inactive

CC mutants); (1)-specific antibodies (Ab); sequences antisense to the (1)  
 CC polynucleotide; agents that downregulate expression of Ese genes or  
 CC antagonists of an Ese binding partner are used to treat diseases  
 CC associated with undesirable endocytosis and resulting changes in cellular  
 CC function. Particularly overexpression of Ese1 is used to block  
 CC clathrin-mediated endocytosis in vivo or in cell cultures, while  
 CC administration of (1) is used to promote endocytosis of selected cells.  
 CC (Ant)agonists of (1) or Ab are used to suppress abnormal proliferation of  
 CC cells that can be stimulated to proliferate by a growth factor receptor;  
 CC and similar compounds (also inactive Ese mutants) can be used to prevent  
 CC viral infection. Endocytosis may also be regulated, in vivo or in cell  
 CC cultures, by forming an Ese-Eps15 complex, then binding dynamn to the  
 CC complex. Generally conditions that can be treated include cancer;  
 CC abnormal cell division or migration; viral infection; or abnormal  
 CC receptor signalling, tissue development or synaptic transmission.  
 CC XX

Sequence 5082 BP; 1363 A; 1253 C; 1391 G; 1075 T; 0 other;

Query Match 99.7%; Score 3630; DB 21; Length 5082;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 1; Gaps 1;

Matches 3641; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 ATGGCTCAGTTCCCAACCTTTGGTGTAGCCTGGATGTGGGCAATACCTGTGAG 60  
 Db 259 ATGGCTCAGTTCCCAACCTTTGGTGTAGCCTGGATGTGGGCAATACCTGTGAG 318  
 QY 61 GAAAGGCGCAAGCATGACAGCAGTTCCTTAGCCTGAAGCCGATAGCGGATTTATTA 120  
 Db 319 GAAAGGCGCAAGCATGACAGCAGTTCCTTAGCCTGAAGCCGATAGCGGATTTATTA 378  
 QY 121 GGTGATCAAGGAGGAACTTTTTCATCTGGTTACCTGAGCTGTCTTAGACAA 180  
 Db 379 GGTGATCAAGGAGGAACTTTTTCATCTGGTTACCTGAGCTGTCTTAGACAA 438  
 QY 181 ATATGGGCGCTAGCGGACATGATAGAGTGAAGATGATCAATGGAATTTTCCATA 240  
 Db 439 ATATGGGCGCTAGCGGACATGATAGAGTGAAGATGATCAATGGAATTTTCCATA 498  
 QY 241 GCCATGAAGCTTATCAAACTGAACTCAAGGATATGAGTCCCTCCACACTTCCCT 300  
 Db 499 GCCATGAAGCTTATCAAACTGAACTCAAGGATATGAGTCCCTCCACACTTCCCT 558  
 QY 301 GTCATGAAGACGAAACCGAGTGTATTTCCAGTCCACAGCATTTGGTATAGAGGAT 360  
 Db 559 GTCATGAAGACGAAACCGAGTGTATTTCCAGTCCACAGCATTTGGTATAGAGGAT 618  
 QY 361 GCTAGCATGCCACCACTACAGCTGTGCTCTGCTCCATAGGCTCATTTCCAGTTGT 420  
 Db 619 GCTAGCATGCCACCACTACAGCTGTGCTCTGCTCCATAGGCTCATTTCCAGTTGT 678  
 QY 421 GGAATGTCTCCACCTTATGATCTTCTGCTCCAGAGAGAGGCTCCCTGGTTAAC 480  
 Db 679 GGAATGTCTCCACCTTATGATCTTCTGCTCCAGAGAGAGGCTCCCTGGTTAAC 738  
 QY 481 GGGGCTCTCCGCTATACAGCTGTGCTGCTGCTGGTGGCATCTCTGACCAATGCGCA 540  
 Db 739 GGGGCTCTCCGCTATACAGCTGTGCTGCTGCTGGTGGCATCTCTGACCAATGCGCA 798  
 QY 541 AAGATTTCTTCTTACAGAGATGTGTCAGAGGTACAAATTAACATTAAGTTACAGAG 600  
 Db 799 AAGATTTCTTCTTACAGAGATGTGTCAGAGGTACAAATTAACATTAAGTTACAGAG 858  
 QY 601 GCACATCATTTGATGTGCGCAGGCGCCCTCCAGACAGAGATGGGCTGTGCTCAGTCA 660  
 Db 859 GCACATCATTTGATGTGCGCAGGCGCCCTCCAGACAGAGATGGGCTGTGCTCAGTCA 918  
 QY 661 TCAAGCTGAAATACAGCAGTATTTCAACAGCCACGACAAATTAAGTGGACACTTA 720  
 Db 919 TCAAGCTGAAATACAGCAGTATTTCAACAGCCACGACAAATTAAGTGGACACTTA 978  
 QY 721 AAGGTCCCGGCAAGAACTATTCTCATGCAATCAAGTTTACCAGGCTCAGCTGGCT 780  
 Db 979 AAGGTCCCGGCAAGAACTATTCTCATGCAATCAAGTTTACCAGGCTCAGCTGGCT 1038





```

Db 3198 GCTAGCTAAAGAGAGTGGCTTCCCGGCCCAAGCCAGCCATTCGCCGAGAAAGTTT 3257
Qy 3001 ATTGGCATGTACATACGAGAGTCTTGAGCAAGAGATTATTAACCTTTTACGAAAGGAGT 3060
Db 3258 ATTGGCATGTACATACGAGAGTCTTGAGCAAGAGATTATTAACCTTTTACGAAAGGAGT 3317
Qy 3061 GTGATTGTGGTTTCCAAAGAAAGATGTGACTGTGAGCCGGAACCGTGGCCGCAACAAGTCC 3120
Db 3318 GTGATTGTGGTTTCCAAAGAAAGATGTGACTGTGAGCCGGAACCGTGGCCGCAACAAGTCC 3377
Qy 3121 GGAGTCTTCCCTCTCTAATCTATGTGAGGCTTTAAAGATTTCAGAGGCTCTTGAAGTCTGGG 3180
Db 3378 GGAGTCTTCCCTCTCTAATCTATGTGAGGCTTTAAAGATTTCAGAGGCTCTTGAAGTCTGGG 3437
Qy 3181 AAAACAGGGAGTTTAGGAAAAAAACCTGAAATTTGCCAGATTATTCCTTCTCAAGCTGCT 3240
Db 3438 AAAACAGGGAGTTTAGGAAAAAAACCTGAAATTTGCCAGATTATTCCTTCTCAAGCTGCT 3497
Qy 3241 ACTGATCCCGAACACTCACTCCCTGCTCTGAGGAGCTGATTTCTGATCCGAAAAAAGAAC 3300
Db 3498 ACTGATCCCGAACACTCACTCCCTGCTCTGAGGAGCTGATTTCTGATCCGAAAAAAGAAC 3557
Qy 3301 CCGAGTGTGATGTGGGAAAGAACTGCAAGCTCGAGGGAAAAAGGCCAGATAGGTTGG 3360
Db 3558 CCGAGTGTGATGTGGGAAAGAACTGCAAGCTCGAGGGAAAAAGGCCAGATAGGTTGG 3617
Qy 3361 TTTCCAGCAATATGTCAAACTTTTAAGCCCGGAAACAAGCAAAATCAACCCCACTGAG 3420
Db 3618 TTTCCAGCAATATGTCAAACTTTTAAGCCCGGAAACAAGCAAAATCAACCCCACTGAG 3677
Qy 3421 CTACCCAAAGCCGAGTGCAGCCAGAGTGTGCAAGTGTGCGGATGTACGATTACACC 3480
Db 3678 CTACCCAAAGCCGAGTGCAGCCAGAGTGTGCAAGTGTGCGGATGTACGATTACACC 3737
Qy 3481 GCCCAGAGCATGACGAACTAGAGCTTCAGCAAAAGGCCAGATCATCAACGCTCCACAAG 3540
Db 3738 GCCCAGAGCATGACGAACTAGAGCTTCAGCAAAAGGCCAGATCATCAACGCTCCACAAG 3797
Qy 3541 GAGGACCCGAGTGTGGAAAGGAAAGTCAAGTGGGCAAGTTGGGCTTTCCCATCAAT 3600
Db 3798 GAGGACCCGAGTGTGGAAAGGAAAGTCAAGTGGGCAAGTTGGGCTTTCCCATCAAT 3857
Qy 3601 TATGTAAAGCTGACCAACAGATGAGACCCAGCCAGCAATGA 3642
Db 3858 TATGTAAAGCTGACCAACAGATGAGACCCAGCCAGCAATGA 3899

RESULT 5
AAZ34572 standard; cDNA; 5195 BP.
ID AAZ34572;
AC AAZ34572;
DT 01-FEB-2000 (first entry)
DE Human SH3D1A cDNA clone 11.
XX
XX SH3D1A gene; human; Down's syndrome; leukemia; cancer;
XX megakaryocytic abnormality; myeloproliferative disorder;
XX platelet disorder; neural disorder; thrombocytopenia;
XX hemotopoietic disorder; cognitive dysfunction; microcephaly;
XX lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy;
XX ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX FH 239..3886
XX FT CDS /*tag= a
XX
XX WO9951062-A2.
XX

```

```

PD 21-OCT-1999.
XX
XX 16-APR-1999; 99MO-US08371.
XX
XX 16-APR-1998; 98US-0082007.
XX
XX (CEDA-) CEDARS SINAI HEALTH SYSTEM.
XX
XX Korenberg JR, Chen X;
XX WPI: 1999-633829/54.
XX P-PSDB; AAY32156.
XX
XX Nucleic acid from the human SH3D1A gene and its products, useful for
XX the diagnosis and treatment of myeloproliferative disorders and
XX leukemia.
XX
XX Claim 2; Fig 10; 99pp; English.
XX
XX This is the nucleotide sequence of full-length cDNA (clone 11)
XX corresponding to a novel human SH3 gene, termed the SH3D1A gene,
XX that contributes to the development of platelets and the
XX pathogenesis of leukemias, both in general and in particular those
XX involving the megakaryocytic lineage. The SH3D1A gene maps to the
XX small candidate region for low platelets on chromosome 21.
XX Sequencing of 5 different sizes of cDNA clone from foetal brain
XX (see AAZ34570-74) suggests that at least 3 isoforms exist. The
XX invention provides methods for the diagnosis and treatment of
XX megakaryocytic abnormality, myeloproliferative disorder, platelet
XX disorder, acute leukemia, neural disorders, thrombocytopenia,
XX platelet disorder on chromosome 21, low platelets in deletion for
XX 21, association of gains in chromosome 21 with leukemias, neural
XX abnormalities, dysfunctions and disorders including brain
XX malformations, and corresponding cognitive dysfunctions,
XX microcephaly, lissencephaly, and colpocephaly. Methods are also
XX provided for: suppressing cells unable to regulate themselves;
XX screening for a somatic alteration in the SH3D1A gene; monitoring
XX the progress and adequacy of a treatment; monitoring tumour risk
XX progress or megakaryocytic abnormality; platelet disorder or leukemia; and
XX haematopoietic disorder, platelet disorder or leukemia; and
XX treatment of a subject (including a prenatal subject) having
XX megakaryocytic abnormality, myeloproliferative disorder,
XX platelet disorder, leukemia or neural disorder using a
XX nucleic acid that expresses SH3D1A or its antisense nucleic acid.
XX
XX Sequence 5195 BP; 1592 A; 1148 C; 1237 G; 1218 T; 0 other;

```

```

Query Match 78.4%; Score 2856; DB 20; Length 5195;
Best Local Similarity 86.8%; Pred. No. 0;
Matches 1167; Conservative 0; Mismatches 475; Indels 6; Gaps 2;

```

```

Qy 1 ATGGCTCAGTTTCCCAACCTTTGGTGTGAGCTGATGCTGGGCATTAAGTGTGAG 60
Db 239 ATGGCTCAGTTTCCCAACCTTTGGTGTGAGCTGATGCTGGGCATTAAGTGTGAG 298
Qy 61 GAAAGGCCAAGCATGACACAGCACTTCTTAGCCTGAAGCCGATAGCGGATTTATTAAT 120
Db 239 GAAAGGCCAAGCATGACACAGCACTTCTTAGCCTGAAGCCAATATCTGATTCATTAAT 358
Qy 121 GGTGATCAAGGAGGAATTTTTTTTCCAAATCTGGCTTAACCTGCTGTTAGCAAA 180
Db 359 GGTGATCAAGGAGGAATTTTTTTTCCAAATCTGGCTTAACCTGCTGTTAGCAAA 418
Qy 181 ATATGGGCGCTAGCGGACATGAATTAAGATGAAGATGATCAAGTGAATTTTCATA 240
Db 419 ATATGGGCGCTAGCGGACATGAATTAAGATGAAGATGATCAAGTGAATTTTCATA 478
Qy 241 GCGATGAAGCTTATCAAACTGAAGCTACAGAGATATCACTCCCTCCACACTTCCCTCT 300
Db 479 GCGATGAAGCTTATCAAACTGAAGCTACAGAGATATCACTCCCTCCACACTTCCCTCT 538
Qy 301 GTCATGAAAGCGAACCGAGTGTATTTCCAGTGCACAGCATTTGATGTGAGAGATT 360

```

Db 539 GTCATGAAACAGCAACAGCTTGTCTATTCTAGGACACACATTTGGTATGGAGATATC 598  
 Qy 361 GCTAGCATGCCACACTCAAGCTGTGTCTCTGTGCCAATGGGCTCCATTCAGTTGTT 420  
 Db 599 GCCAGATGCCACCGCTTACAGCTGTGTCTCCAGTGCCTCAATGGGATCCATTCAGTTGTT 658  
 Qy 421 GGAATGTCTCCACCTTACTATCTTCTGTCCCTCCAGCAGCAGTGTGCTTCCCTGGCTAAC 480  
 Db 659 GGAATGTCTCCAAACCTTATCTTCTGTCCCAAGCAGCAGTGTGCTTCCCTGGCTAAC 718  
 Qy 481 GGGGCTCTCCCGCTCATACAGCTCTGTGCTGGTGGCCATCTCCAGCAGCATGGCCA 540  
 Db 719 GGGGCTCTCCCGCTGTATACAACTCTGTGCTCATTTGCTCATCTCCAGCAGCATTTGCCA 778  
 Qy 541 AAGAGTTCTTCTTCCAGCATCTGTGTCCAGGCTCAATTAACACTAAGTTACAGAA 600  
 Db 779 AAGAGTTCTTCTTGTAGATCTGTGTCCAGGCTCAAACTAAGCTAATATACAAAG 838  
 Qy 601 GCACATCATTTGATGTGCGCAGCGCCCTCCAGCAGCAGAAATGGGCTGTGCTCATGTCA 660  
 Db 839 GCACATCATTTGATGTGCGCAGGTGCCACAGTGGCAGAGTGGGCTGTCTCATGTCA 898  
 Qy 661 TCAGGCTGAAATACAGGAGATTATCAACAGCAGCAAACTATGATGAGACCTTA 720  
 Db 899 TCAGGCTGAAATACAGGCAATTTATCAATAGTCATGACAAAACATGTAGTGAACCTTA 958  
 Qy 721 ACAGGTCCCAAGCAGAACTATCTCATGCAATCAAGTTTACCCAGGCTCAGTGTCT 780  
 Db 959 ACAGGTCCCAAGCAGAACTATCTTATGCAGTCAAGTTTACCAAGCTCAGCTGGCT 1018  
 Qy 781 TCATATGGAATCTTTCTGACATTTGATCAAGATGGAATCTCAGTGCAGAAATTTATC 840  
 Db 1019 TCATATGGAATCTTTCTGACATTTGATCAAGATGGAATCTTACAGCAGAGAAATTTATC 1078  
 Qy 841 CTAGCATGACCTTAATGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
 Db 1079 CTGCAATGCACTTATGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1138  
 Qy 901 CCAGAAATACATCCCTCTTCTTCCAGAGAGTTGCTGCGGAGTGGATGTCCTCATTA 960  
 Db 1139 CCAGAAATACATTCACCTTCTTCCAGAGAGTTGCTGCGGAGTGGATGTCCTCATTA 1198  
 Qy 961 AGCTCTTCTTCTGTGATCAGAGGCTGCTGAGAGCCGCTGCTGAGAGATGACAGCAG 1020  
 Db 1199 AGCTCAACATCTGTAGATCAGAGGCTTACCAAGAGAACTTTTGAATGAACAACA 1258  
 Qy 1021 C---CAGAGAGAACTGCTGTGACATTTGAATGAATGAAGGGAGAACTTCAGACGA 1077  
 Db 1259 CAATTAAGAAAGAAATTTACCTGTAACTTTGAAGATGAAGCGGAGAACTTTGAACGT 1318  
 Qy 1078 GGCAGTGTGAGCTGAGAGAGCGCCGCAAGCGCTCTTGAGAGCAGAGCGCAAGAGCAG 1137  
 Db 1319 GGCAGTGTGAGCTGAGAGAGCGAAAGCGAAAGCTCTCTGGAACAGAGCGCAAGAGCAG 1378  
 Qy 1138 GAGCGTGTGCTGAGTGAAGCGCGCGAGCAGAGAGAGAAAGCGGAGCGCCAGAGAG 1197  
 Db 1379 GAGCGCTGTGCGCAGCTGAGCGCGGAGCAGAGAGAGAGAGAGCGTGAAGCGCAGAGAG 1438  
 Qy 1198 CAGGAGCGCAAGCGGAGCTGAGCTGAGAGAGCAGCAGTGAAGAGAGCGGAGCTGAGAG 1257  
 Db 1439 CAGAGCGCAAGAGAGCACTGAGAACTGAGAAAGCAACTGAGAAAGAGCGGAGCTGAGAA 1498  
 Qy 1258 CGCAGCAGAGAGAGAGAGAGAGAGAGATGAGAGAGCGCGAGCGCCAGAAACGGGAA 1317  
 Db 1499 CGCAGCAGAGAGAGAGAGAGAGAGAGAAATTTGAGAGCGCAGAGGCTGCCAAAACGGGAA 1558  
 Qy 1318 CTGGAAGAGCAGCAGCACTTGAATGGAAACGGAACTGGAACCGGAACTCTTGAATCAG 1377  
 Db 1559 CTGGAAGAGCAGCAGCACTTGAATGGAAACGGAACTGGAACCGGAACTCTTGAATCAG 1618  
 Qy 1378 AGGAACAAG 1437  
 Db 1619 AGGAACAAG 1678

Qy 1438 GAGTTAGAGCTCTGAATGACAAAAACATACAGTAGAGGAAACTTCAGATATCAGG 1497  
 Db 1679 GAATTAGAGCTCTTAATGATTAATAACATCACTAGAAAGGAAACTTCAAGATATCAGA 1738  
 Qy 1498 TGTGAGCTGGCAACCCAGAGCGAAAGAAATTTGAGAGCAACGAAGTCTTAGAGCTAAGA 1557  
 Db 1739 TGTGAGTTGACCAACCAGAGCAAGAAATTTGAGAGCAACGAAGTCTTAGAGCTAAGA 1798  
 Qy 1558 ATTGCTGAAATCACCCACTTACAGCAGCAGTGGAGAGATCTCAGCAAAATGCTTGAAGA 1617  
 Db 1799 ATTGCCAATCACCCACTTACAGCAGCAGTGGAGAGATCTCAGCAAAATGCTTGAAGA 1858  
 Qy 1618 CTATTCAGAGAAACAGATATCTCAGTGACAGTAAACAGTTCACAGCAAAAGTTTG 1677  
 Db 1859 CTATTCAGAGAAACAGATATCTCAGTGACCAATTAATAAGTTTCAGAGAAAGTTTG 1918  
 Qy 1678 CATAGACATCGCTTCTTACCCTCAAAAGAGCTTGGAAACCAAGAGCTGGCCGCGCAG 1737  
 Db 1919 CACAGAGATTCATTTGTATACATTTAAAGAGCTTGAAGCAAAAGAGCTTGGCAG 1978  
 Qy 1738 CAGCTCCGAGCAGCTGAGCAGAGGTGAGAGAGAGACAGTTCAAAGCTGACAGAGATT 1797  
 Db 1979 CACCTAGAGAGCAACTGATGAGATGAGAGAGAGAACTAGATCAAACTACAGAGATT 2038  
 Qy 1798 GATGTTTCAACCAACAGCTGAGAGAACTGAGAGAGATACATGCAAAACAGCAACTTCAG 1857  
 Db 2039 GATATTTCAATATCATGCTGAAGAGAACTAAGAGAAATACAACTAAGCAAACTCCAG 2098  
 Qy 1858 AAGCAGAGTGTCCCTGGAGAGCAGCCGCACTGAGACGAAAGAGCAGAGAGAGAGAGCTG 1917  
 Db 2099 AAGCAAAAGTCAATGAGAGGCTGAAAGCACTGAAACAGAAACAGAAACAGAAAGATCAT 2158  
 Qy 1918 GAGTTAGAGAAACAAAGAGAGAGCTCAGAGAGAGTTCAGGAAAGAGAGCAAGTGG 1977  
 Db 2159 GAATTAAGAAACAAAG 2218  
 Qy 1978 CTGAGAGATGTGAGCAGAGAG--GAGCAGCAGCGCCCGGAAACCCAGAGAGAGAG 2034  
 Db 2219 CTGAGAGATGTGAGCAG 2278  
 Qy 2035 AGACTGAAGAGAGAGAGAGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2094  
 Db 2279 AAACCTGAAAGAGAGAGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2338  
 Qy 2095 ATGCAAGCAAGCAGAGTGGGCTTTTCATCCGATCAGAGAGCAGTGAAGTGGCCACC 2154  
 Db 2339 GCACAG 2398  
 Qy 2155 CAGCAGCCTGTGCTTACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2214  
 Db 2399 CAGCAGCCTGTGCTTACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2458  
 Qy 2215 AAAGTGTATTTACCGAGAGCGCTGTACCCCTTTGAATCCGAAGTCAAGATGATACC 2274  
 Db 2459 AAAGTGTGTATTTACCGAGAGCGCTGTACCCCTTTGAATCCGAAGTCAAGATGATACC 2518  
 Qy 2275 ATCCAGCAGAGAGATATAGTATGATGTGATGAAGCCAGACTGAGAGAGCAGAGATGGCTT 2334  
 Db 2519 ATCCAGCAGAGAGATATAGTATGATGTGATGAAGCCAGACTGAGAGAGCAGAGATGGCTT 2578  
 Qy 2335 GAGAGAGAGCTGAAAG 2394  
 Db 2579 GAGAGAGAGATTAAG 2638  
 Qy 2335 GAAATAGAGTTCCCACTCCAGCCAAACAGTGAACGATGTGACATGTGCTGCTGCTGCTG 2454  
 Db 2639 GAAATAGAGTTCCCACTCCAGCCAAACAGTGAACGATGTGACATGTGCTGCTGCTGCTG 2698  
 Qy 2455 AAACGTGCTGCTGAG 2514  
 Db 2699 AAACGTGCTGCTGAG 2758

QY 2515 ACCCCCAACAACCTGGGAGAGCTTCACTTCCAGTGGGCCAGAGCTCAACAGGAAGCA 2574  
 Db 2759 ACCCTTAATTAAGTGGGCGGAGCTTCACTTCCAGTGGGCCAGAGCTCAACAGGAAGCA 2818  
 QY 2575 GAAACGACAACTGGGAGATACGTGGGGGGCTCAGCCTTCTCTGACCGTACTAGTGTG 2634  
 Db 2819 GAAACGATTAAGTGGGAGTGTGGGAGCGCCAGCCTCTCTCACTGCTTCAAGTGGCG 2878  
 QY 2635 CAGTTACGGCAGATGATGAGCTTTTACCCAGCCAGCAGCCACTGAGCTTCTCTCC 2694  
 Db 2879 CAGTTAAGGAGAGGCTCCGCTTTACTTCAAGCCAGCCAGCTGAGCTTCTCTCC 2938  
 QY 2695 GTCTGGGGCAGGAGTGAAGAGGCTCAAGCCGAGCCCTGATTCCTTGGAGA 2754  
 Db 2939 GTGCTAGGCGAGGAGTGAAGAGGCTCAAGCTCAAGCCCTTAATCTTGGAGA 2998  
 QY 2755 GCCAAAAAGACACCACTTAAATTTTACAAAAGTGAAGCTTCACTGCTTGGAG 2814  
 Db 2999 GCCAAAAAGACACCACTTAAATTTTACAAAAGTGAAGCTTCACTGCTTGGAG 3058  
 QY 2815 CAAGACATGAGTGGTGTGGAGAGTCAAGGTCAGAGGAGGTTCCCAAGCTTAC 2874  
 Db 3059 CAAGACATGAGTGGTGTGGAGAGTCAAGGTCAGAGGAGGTTCCCAAGCTTAC 3118  
 QY 2875 GTGAACTCATTTCAAGGCGCCGTAAGGAAATCCAGAGCTGATCTGAGCTTCTGAA 2934  
 Db 3119 GTGAACTCATTTCAAGGCGCCGTAAGAGTCAAGAGCTGATCTGAGCTTCTGAGAG 3178  
 QY 2935 AGTCTGCTGCTTCAAGAGAGTGTCCCGCGCCAGAGCAGCACTTCCCGAGAA 2994  
 Db 3179 AGTCTGCTGCTTCAAGAGAGTGTCCCGCGCCAGAGCAGCACTTCCCGAGAA 3238  
 QY 2995 GAGTTATTCATGTAACATACGAGAGTTCAGAGCAGAGATTTAATCTTACGAA 3054  
 Db 3239 GAGTTATTCATGTAACATACGAGAGTTCAGAGCAGAGATTTAATCTTACGAA 3298  
 QY 3055 GGGGATGTGATTTGTTTACCAAGAAAGTGTGATCTGTGAGCGGAGAGGTTGGCG 3114  
 Db 3299 GGGGATGTGATTTGTTTACCAAGAAAGTGTGATCTGTGAGCGGAGAGGTTGGCG 3358  
 QY 3115 AAGTCCGAGCTTCCCTTCACTATGAGGCTTAAAGATTCAGAGGCTTGGAACT 3174  
 Db 3359 AAGTCCGAGCTTCCCTTCACTATGAGGCTTAAAGATTCAGAGGCTTGGAACT 3418  
 QY 3175 GCTGGGAAAAAGGAGGTTTGAAGAAAAAACTGAAATTCGAGGTTATTCCTTAC 3234  
 Db 3419 GCTGGGAAAAAGGAGGTTTGAAGAAAAAACTGAAATTCGAGGTTATTCCTTAC 3478  
 QY 3235 GCTGCTACTGCTCCGGAACAACCTGAGCTGCTTGGGAGCTGATTCGATCCGAAA 3294  
 Db 3479 ACCGCAACCGGCGGAGAGCTGCTGCTTGGGAGCTGATTCGATTCGAAA 3538  
 QY 3295 AAGAACCAAGTGTGATGTGGGAAAGAACTGCAAGCTGAGGAAAAAGCCAGATA 3354  
 Db 3539 AAGAACCAAGTGTGATGTGGGAAAGAACTGCAAGCTGAGGAAAAAGCCAGATA 3598  
 QY 3355 GGGTGTTCCTCAGAAATTAATGCAACTTAAAGCCCGGAAACCAACAATTCACCCA 3414  
 Db 3599 GGGTGTTCCTCAGAAATTAATGCAACTTAAAGCCCGGAAACCAACAATTCACCCA 3658  
 QY 3415 ACTGAGTACCAAGACCGGAGCTGAGCAGAGTGTGAGGAGTGTGAGTGTAGAT 3474  
 Db 3659 ACTGAGTACCAAGACCGGAGCTGAGCAGAGTGTGAGGAGTGTGAGTGTAGAT 3718  
 QY 3475 TACACCGCCAGAAAGATGACGAATCTGCTTCAAGAAAGCCAGATCAACGTTCTC 3534  
 Db 3719 TACACCGCCAGAAAGATGACGAATCTGCTTCAAGAAAGCCAGATCAACGTTCTC 3778  
 QY 3535 AACAGAGGAGACCGGAGCTGTGAAAGAGAGTGTGAGGAGTGTGAGTGTGAGTGTGAG 3594  
 Db 3779 AACAGAGGAGACCGGAGCTGTGAAAGAGAGTGTGAGGAGTGTGAGTGTGAGTGTGAG 3838  
 QY 3595 TCCAATTATGTAAAGCTGACCAAGACATGAGCCCAAGCCAGCAATGA 3642

Db 3839 TCCAATTATGTGAAGCTGACCAAGACATGAGCCCAAGCAGCAATGA 3886  
 RESULT 6  
 AA234571  
 ID AA234571 standard; cDNA; 5458 BP.  
 XX  
 AC AA234571;  
 XX  
 XX 01-FEB-2000 (first entry)  
 DT  
 DE Human SH3D1A cDNA clone 21.  
 XX  
 DE SH3D1A gene; human; Down's syndrome; leukemia; cancer;  
 KW megakaryocytic abnormality; myeloproliferative disorder;  
 KW platelet disorder; neural disorder; thrombocytopenia;  
 KW haematopoietic disorder; cognitive dysfunction; microcephaly;  
 KW lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy;  
 KW ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 267..392  
 FT /\*tag= a  
 PN MO9953062-A2.  
 XX  
 PD 21-OCT-1999.  
 XX  
 PF 16-APR-1999; 99NC-U508371.  
 XX  
 PR 16-APR-1998; 98US-0082007.  
 XX  
 PA (CEDA-) CEDARS SINAI HEALTH SYSTEM.  
 XX  
 PI Korenberg JR, Chen X;  
 XX  
 DR WP1: 1999-633829/54.  
 XX  
 DR P-PSDB; AAY32155.  
 XX  
 PT Nucleic acid from the human SH3D1A gene and its products, useful for  
 PT the diagnosis and treatment of myeloproliferative disorders and  
 PT leukemia -  
 XX  
 PS Claim 2; Fig 8; 99pp; English.  
 XX  
 CC This is the nucleotide sequence of full-length cDNA (clone 21)  
 CC corresponding to a novel human SH3 gene, termed the SH3D1A gene,  
 CC that contributes to the development of platelets and the  
 CC pathogenesis of leukemias, both in general and in particular those  
 CC involving the megakaryocytic lineage. The SH3D1A gene maps to the  
 CC small candidate region for low platelets on chromosome 21.  
 CC Sequencing of 5 different sizes of cDNA clone from foetal brain  
 CC (see AA234570-74) suggests that at least 3 isoforms exist. The  
 CC invention provides methods for the diagnosis and treatment of  
 CC megakaryocytic abnormality, myeloproliferative disorder, platelet  
 CC disorder, acute leukemia, neural disorders, thrombocytopenia,  
 CC platelet disorder on chromosome 21, low platelets in deletion for  
 CC 21, association of gains in chromosome 21 with leukemias, neural  
 CC abnormalities, dysfunctions and disorders including brain  
 CC malformations and corresponding cognitive dysfunctions,  
 CC microcephaly, lissencephaly, and colpocephaly. Methods are also  
 CC provided for a somatic alteration in the SH3D1A gene; monitoring  
 CC screening for a somatic alteration in the SH3D1A gene; monitoring  
 CC the progress and adequacy of a treatment; monitoring tumour risk  
 CC progress or megakaryocytic abnormality, myeloproliferative disorder,  
 CC haematopoietic disorder, platelet disorder or leukemia; and  
 CC treatment of a subject (including a prenatal subject) having  
 CC megakaryocytic abnormality, myeloproliferative disorder,  
 CC platelet disorder, leukemia or neural disorder using a  
 CC nucleic acid that expresses SH3D1A or its antisense nucleic acid.

XX Sequence 5458 BP, 1671 A; 1201 C; 1307 G; 1279 T; 0 other;  
 SQ Query Match 77.7%; Score 2831; DB 20; Length 5458;  
 Best Local Similarity 86.5%; Pred. No. 0;  
 Matches 3167; Conservative 0; Mismatches 475; Indels 21; Gaps 3;

QY 1 ATGGCTGACGTTTCCACACCTTTGGTGTAGGCTGTGATGTGGCCATTAAGTGTGAG 60  
 DB ATGGCTGACGTTTCCACACCTTTGGTGTAGGCTGTGATGTGGCCATTAAGTGTGAG 326  
 QY 61 GAAAGGCGAGGATGACGAGGCTTCCCTTACGCTGAAGCCGATACCGGATTTATCT 120  
 DB GAAAGGCGAGGATGACGAGGCTTCCCTTACGCTGAAGCCGATACCGGATTTATCT 386  
 QY 121 GGTGATCAAGCGAGAACTTTTTCATCTGGTTACCTCAGCCTGTCTTACGACAA 180  
 DB GGTGATCAAGCGAGAACTTTTTCATCTGGTTACCTCAGCCTGTCTTACGACAA 446  
 QY 181 ATATGGGCGGTAGCGGACATGAATTAACGATGAAGATGAGATGAAATTTTCATA 240  
 DB ATATGGGCGGTAGCGGACATGAATTAACGATGAAGATGAGATGAAATTTTCATA 506  
 QY 241 GCGATGAAGTTTCAAACTGAAGCTTAAGGATATCAGTCCCTTCCACACTTCCCT 300  
 DB GCGATGAAGTTTCAAACTGAAGCTTAAGGATATCAGTCCCTTCCACACTTCCCT 566  
 QY 301 GTCATGAAACAGCAACAGTGGCTATTTCCAGTGCACAGCATTTTGTATGAGGAGTT 360  
 DB GTCATGAAACAGCAACAGTGGCTATTTCCAGTGCACAGCATTTTGTATGAGGAGTT 626  
 QY 361 GGTAGCATGCGCACCTACAGAGTGTGCTCTGTGCGCAATGGGCTCATTCAGTTGTT 420  
 DB GGTAGCATGCGCACCTACAGAGTGTGCTCTGTGCGCAATGGGCTCATTCAGTTGTT 686  
 QY 421 GGAATGCTCCACCTTATGATCTTCTGTCCCTCCAGACAGATGCTCTCCCTGCTTAC 480  
 DB GGAATGCTCCACCTTATGATCTTCTGTCCCTCCAGACAGATGCTCTCCCTGCTTAC 746  
 QY 481 GGGGCTCTCCCTGCTATACAGCTTGCCTGCTTGGCGATCTGCGACATGAGCA 540  
 DB GGGGCTCTCCCTGCTATACAGCTTGCCTGCTTGGCGATCTGCGACATGAGCA 806  
 QY 541 AAGAGTTCTTCTTACAGCAGATCTGCTCAGGGTCACTAATTAACAATAAGTTCAGAA 600  
 DB AAGAGTTCTTCTTATGATCTGCTCAGGGTCACTAATTAACAATAAGTTCAGAA 866  
 QY 601 GCACATCATTCATGTCGCGCAGCGCCCTCCAGCAGCAAGATGGGCTGTCTCAGTCA 660  
 DB GCACATCATTCATGTCGCGCAGCGCCCTCCAGCAGCAAGATGGGCTGTCTCAGTCA 926  
 QY 661 TCAAGGCTGAATATACAGGAGTATTAACAAGCAGCAAGAACTATGATGAGCACTTA 720  
 DB TCAAGGCTGAATATACAGGAGTATTAACAAGCAGCAAGAACTATGATGAGCACTTA 986  
 QY 721 ACAGGTCCTCCAGGCAAGAACTATTCATGATCAAGATTACCCAGGCTCAGTGGCT 780  
 DB ACAGGTCCTCCAGGCAAGAACTATTCATGATCAAGATTACCCAGGCTCAGTGGCT 1046  
 QY 781 TCAATATGAAATTTTCTGATTTGATCAAGATGAAACTACTGCAAGAAATTTATC 840  
 DB TCAATATGAAATTTTCTGATTTGATCAAGATGAAACTACTGCAAGAAATTTATC 1106  
 QY 841 CTAGCTATGCACTTAATGATGTTGGCATGTGATCAGCACTGCGCGCCGCTGCTGCT 900  
 DB CTAGCTATGCACTTAATGATGTTGGCATGTGATCAGCACTGCGCGCCGCTGCTGCT 1166  
 QY 901 CCAAGATATACCTCTCTCTCTTCAAGAGTTCGCTCCGCGAGTGGAGTGTCCGTCATA 960  
 DB CCAAGATATACCTCTCTCTCTTCAAGAGTTCGCTCCGCGAGTGGAGTGTCCGTCATA 1226  
 QY 961 AGCTCTTCTTCTGATCAAGAGCTGCTGAGAGCCGCTGCTCAAGATGAGCAGAG 1020

DB 1227 AGCTCAACATCTGTATGATCAAGAGCTTACAGAGAACCCAGTTTATGAAATGAACAA 1286  
 QY 1021 C---CAGAGAAAGAACTGCTGTGACATTTGAAATGAAGAGCGGAGAACTTGCAGCA 1077  
 DB 1287 CAATTAGAAAAGAAATTAACCTGTAACTTTAAGATTAAGAACCGGAGAACTTTGAACT 1346  
 QY 1078 GGCAGTGTGAGCTGTGAGAGCGCCGCGCAAGCGCTCTTGGAGCAGCAGCGCAAGAGCAG 1137  
 DB 1347 GGCAGCTGTGAGCTGTGAGAGCGCCGCGCAAGCGCTCTTGGAGCAGCAGCGCAAGAGCAG 1406  
 QY 1138 GAGCGGTTGGCTCAGCTGAGACCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1197  
 DB 1407 GAGCGCTGGCTCAGCTGAGACCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1466  
 QY 1198 CAGAGGCGCAAGCGGCACTGAGAGTGAAGAGAGCTGTGAGAGAGAGAGAGAGAGAGAG 1257  
 DB 1467 CAGAGGCGCAAGCGGCACTGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1526  
 QY 1258 CGGAGCGAG 1317  
 DB 1527 CGGAG 1586  
 QY 1318 CTGGAAGGAG 1377  
 DB 1587 CTGGAAGGAG 1646  
 QY 1378 AGAACAAG 1437  
 DB 1647 AGAACAAG 1706  
 QY 1438 GAGTTAGAGCTCTGATGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1497  
 DB 1707 GAGTTAGAGAGCTCTGATGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1766  
 QY 1498 TGTGACCTGCGCAACCGAG 1557  
 DB 1767 TGTGACCTGCGCAACCGAG 1826  
 QY 1558 ATTGCTGAATCAACCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617  
 DB 1827 ATTGCTGAATCAACCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1886  
 QY 1618 CTATTTCCAG 1677  
 DB 1887 CTATTTCCAG 1946  
 QY 1678 CATGAGAGCTGCTTTTACCTTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1737  
 DB 1947 CATGAGAGCTGCTTTTACCTTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2006  
 QY 1738 CAGCTCCGAG 1797  
 DB 2007 CAGCTCCGAG 2066  
 QY 1798 GATGTTTCAACCAAGCTGAG 1857  
 DB 2067 GATGTTTCAACCAAGCTGAG 2126  
 QY 1858 AAGCAGAGTCCCTGAG 1917  
 DB 2127 AAGCAGAGTCCCTGAG 2186  
 QY 1918 GAGTTAG 1977  
 DB 2187 GAGTTAG 2246  
 QY 1978 CTGAGAGTGTGAG 2034  
 DB 2247 CTGAGAGTGTGAG 2094  
 QY 2035 AGACTGAG 2094  
 DB 2307 AGACTGAG 2366

QY 2095 ATGCAAGACAGAGAGTGGCTTTTCATCCGATCAGAGACCGACTAGCTGGCCACC 2154  
 Db 2367 GCACAAAGACAGAGTGGCTTTTCATCAACACCAAGAACAGCTAGGCACTGCTC 2426  
 QY 2155 CAGGACCTGTGTCTACACAGAGAAAGCCCGCTTACATTTCTGACAGAGAGGTGA 2214  
 Db 2427 CAGGACCTGTGTCTACACAGAGAAAGGTCCATTACATTTCTGACAGAGAAATGA 2486  
 QY 2215 AAAGGTATATTAACGAGAGGTGTACCCCTTTGAATCCAGAACTACAGATGATCA 2274  
 Db 2487 AAAGGTATATTAACGAGAGGTGTACCCCTTTGAATCCAGAACTATGATGATCACT 2546  
 QY 2275 ATCCAGCCAGAGATATATGAT-----GTTGATGAAGCCAGCTGGA 2319  
 Db 2547 ATCCAGCCAGAGATATATGAT-----GTTGATGAAGCCAGCTGGA 2606  
 QY 2320 GACCCAGATGCTGTGAGAGAGCTGAAGAGGAGACGGATGTTCCCTGCAACTAT 2379  
 Db 2607 GACCCAGATGCTGTGAGAGAGATTAAGAGAGAGACGGATGTTCCCTGCAACTAT 2666  
 QY 2380 GCAGAAAGATTCAGAGAAATGAGTTCCACTCCAGCCAAACCACTGACCGATCTGACA 2439  
 Db 2667 GCAGAGAAATTCAGAGAAATGAGTTCCCGCTCCAGTAAACAGTGACTGATCAACA 2726  
 QY 2440 TCTGCCCCCTGCCCCCAAACTGCTCTGCTGAGAACCCCTGCTCTTGGCAGTGAACCT 2499  
 Db 2727 TCTGCCCCCTGCCCCCAAACTGCTCTGCTGAGAACCCCGCTTGGCAGTGAACCTCT 2786  
 QY 2500 TCTGAGCCCTCCAGCAACCCCAACAACTGGGAGAGCTTCAAGTTCAGCTGAGCCAGC 2559  
 Db 2787 TCAAGCCCTCCAGCAACCCCAACAACTGGGAGAGCTTCAAGTTCAGCTGAGCCAGC 2846  
 QY 2560 TCAAGCCAGAGCCAGAGACGACAACTGGGATAGTGGGCGGCTAGCTTCTGACACC 2619  
 Db 2847 ACAGATGAGAGAAACAGAGAAAGGATTAACCTGGATGATGGGCAACCCAGCTCTTCAC 2906  
 QY 2620 GTACCTAGTGTGGCCAGATTACGAGAGATCAGCCCTTACCCAGCCAGCCAGCTGAGC 2679  
 Db 2907 GTTCCAGTGTGGCCAGATTAAAGCAGAGATCCGCTTTTCTCAGCCAGCCAGCTGAGC 2966  
 QY 2680 TCTTCCCATCTCCCGTCTGAGGCGAGGATGAAGAGGTGAGAGGCTCAACGCGCAGCC 2739  
 Db 2967 TCTTCCCGTCTCCCGTCTGAGGCGAGGATGAAGAGGTGAGAGGCTCAACGCGCAGCC 3026  
 QY 2740 CTGATCTCTGAGAGGCGCAAAAAGACACCACTTAATTTTAAACAAAGTGCATC 2799  
 Db 3027 CTATATCTTGGAGAGCGCAAAAAGACACCACTTAATTTTAAACAAAGTGTGATC 3086  
 QY 2800 ACCGTTCTGAGAACAGAGACATGTGTGTTGAGAGAGTTCAAGTCAAGAGGCTTGG 2859  
 Db 3087 ACCGTTCTGAGAACAGAGACATGTGTGTTGAGAGAGTTCAAGTCAAGAGGCTTGG 3146  
 QY 2860 TTCCCAAGTCTTAAGTGAATCTATTTCAAGGCGCTTAAGAAATTCACAAAGATCGAT 2919  
 Db 3147 TTCCCAAGTCTTAAGTGAATCTATTTCAAGGCGCTTAAGAAAGTCTTAACACATGAT 3206  
 QY 2920 ACTGCCCCCTGTAAGTCTGCTAGTCTAAAGAGTGTGCTCCCGGCGCCAGAGCCA 2979  
 Db 3207 TCTGTTCTTCAAGAGTCTGCTAGTCTAAAGAGTGTGCTCCCGGCGCCAGAGCCA 3266  
 QY 2980 GCAATTCGCGAGAGAGTATTTGCAATGATCAATACAGAGATTTCTGAGCAAGAGAT 3039  
 Db 3267 GTGTTTCGCGAGAGAGTATTTGCAATGATCAATACAGAGATTTCTGAGCAAGAGAT 3326  
 QY 3040 TTAACCTTTTCAAGAGAGGATGTGATTTGTTTCAAGAAAGATGTTGCTGTGAGC 3099  
 Db 3327 TTAACCTTTTCAAGAGAGGATGTGATTTGTTTCAAGAAAGATGTTGCTGTGAGC 3386  
 QY 3100 GGAACGGTGGCGCAAGATCCGAGTCTTCCCTTCAACTATGTGAGGCTTAAAGATTCA 3159  
 Db 3387 GGAACGGTGGCGCAAGATCCGAGTCTTCCCTTCAACTATGTGAGGCTTAAAGATTCA 3446

QY 3160 GAGGCTCTGAACTGTGGGAAACAGGAGATTAGAGAAAAAACTGAATTTGCCAG 3219  
 Db 3447 GAGGCTCTGAACTGTGGGAAACAGGAGATTAGAGAAAAAACTGAATTTGCCAG 3506  
 QY 3220 GTTATGCTTCTCAACGCTGTACTGTGTCCGAACTACCTGCTCTTGGGACGCTG 3279  
 Db 3507 GTTATGCTCTATACACCGCCACCGCCCGAGCAGCTCACTCTCGCCCTGTGCTG 3566  
 QY 3280 ATTGTGATCCGGAAGAAAGAACCCAGTGTGATGTGGAGAGAGAACTGCAAGCTCGAGG 3339  
 Db 3567 ATTGTGATCCGGAAGAAAGAACCCAGTGTGATGTGGAGAGAGCTGCAAGCTCGAG 3626  
 QY 3340 AAAAGCCAGATAGGAGTGTTCAGCAAAATATATGCAAACTTGAACCCCGGAGCA 3399  
 Db 3627 AAAAGCCAGATAGGAGTGTTCAGCAAAATATATGTAAGCTTGAAGCCCTGGAGC 3686  
 QY 3400 AGCAAAATCAACCCCACTGAGCTACCAAGACCGCAGTGCAGCAGAGTGTGCCAGTG 3459  
 Db 3687 AGCAAAATCACTCCACAGAGCCACCTTAAGTCAACAGCATTAAGCCGAGTGTCCAGTG 3746  
 QY 3460 ATGGGATGTACGATTAACCGCCAGAAAGATGACGAACTAGCTTCAAGCAAGCCAG 3519  
 Db 3747 ATGGGATGTACGATTAACCGCCAGAAATGACGATGAGTGTGCTTCAACAGGCGCAG 3806  
 QY 3520 ATCATCAACGCTCTCAACAGAGAGACCCGAGTGTGAGAGAGAGTCACTGGGCAA 3579  
 Db 3807 ATCATCAACGCTCTCAACAGAGAGACCTTGAAGTGTGAGAGAGAGTCAATGAGCAA 3866  
 QY 3580 GTTGGGCTCTTCCATCACTTATGTAAAGTGAACCAACAGATGAGCCAGCCAGCAA 3639  
 Db 3867 GTTGGGCTCTTCCATCACTTATGTAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 3926  
 QY 3640 TGA 3642  
 Db 3927 TGA 3929

RESULT 7  
 AAS84763  
 ID AAS84763 standard; cDNA; 7435 BP.  
 AC AAS84763;  
 DT 13-FEB-2002 (first entry)  
 DE DNA encoding novel human diagnostic protein #20567.  
 KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KM food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 OS Homo sapiens.  
 PN W0200175067-A2.  
 PD 11-OCT-2001.  
 PF 30-MAR-2001; 2001WO-US08631.  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 PA (HYSE-) HYSEQ INC.  
 PI Drmanac RT, Liu C, Tang YT;  
 PI WPI; 2001-639362/73.  
 DR P-PSDB; ABG20576.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX

PS Claim 1; SEQ ID No 20567; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 7435 BP; 2190 A; 1690 C; 1829 G; 1726 T; 0 other;

Query Match 68.4%; Score 2491.6; DB 23; Length 7435;  
Best Local Similarity 84.2%; Pred. No. 0;  
Matches 3112; Conservative 0; Mismatches 529; Indels 53; Gaps 25;

QY 1 ATGGCTCACTTTCCACACCTTTGGTGTAGCTGTGATGTCGTGAGCCATACTGTGAG 60  
DB 261 ATGGCTCACTTTCCAAACCTTTGGTGTAGCTGTGATGTCGTGAGCCATACTGTGAG 322  
QY 61 GAAAGGCCAACAGTACAGAGATTCCTTACCTGAGCCATGCGGAGTTTATTACT 120  
DB 323 GAAAGGCCAACAGTACAGAGATTCCTTAAACCAATATCTGGATTCCTTACT 382  
QY 121 GGTGATCAAGCAGAGAACTTTTTCATCTGGGTTACTCAGCTGTCTTACACAA 180  
DB 383 GGTGATCAAGTACAGAACTTTTTCATCTGGGTTACTCAGCTGTCTTACACAG 442  
QY 181 ATATGGGCGCTAGCGAGCATGAATAAGATGAGAGATGATCAAGTGAATTTTCATA 240  
DB 443 ATATGGGCGCTAGCTAGCATGAATAAGATGAGAGATGATCAAGTGAATTTTCATA 502  
QY 241 GGCATGAAGCTTATCAAACTGAAGCTACAGATATCAGCTCCCTCCACACTTCCCT 300  
DB 503 GGTATGAACCTTATCAAACTGAAGCTACAGATATCAGCTCCCTCCACACTTCCCT 562  
QY 301 GTCATGAACAGCAACAGTGTGCTATTTTCAGTGACAGCACTTTGGTATAGAGAGATT 360  
DB 563 GTCATGAACAGCAACAGTGTGCTATTTTCAGTGACAGCACTTTGGTATAGAGAGATT 622  
QY 361 GGTAGCATGCACTCACTCACTGTGCTGTGCTGTGCAATGGAGCTTCATT-CCAGTTGT 419  
DB 623 GCGAGATGCACTCACTGTGCTGTGCTGTGCTGTGCAATGGAGCTTCATTCCAGTTGT 682  
QY 420 TGGATGTCTCACTTGTATGTTTCTCTCCCTCAGAGAGAGTCTCCCTGGCTTAA 479  
DB 683 TGGATGTCTCACTTGTATGTTTCTCTCCCTCAGAGAGTCTGGCCCCCTGGCTTAA 742  
QY 480 CCGGGCTCTCCGTCATACAGCTCTGCTGCTGTTGGCATCTCCGAGAGCAGATGGCC 539  
DB 743 CCGGGCTCTCCGTCATACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 802  
QY 540 AAGAGTTCTCTTCTCAGCAGATCTGTGCTCAGAGTCACAATTAAACATTAATTACAGAA 599  
DB 803 AAGAGTTCTCTTCTTAGTAGATCTGTGCTCAGAGTCACAATTAAACATTAATTACAGAA 862  
QY 600 GGCACATATTTGATGTGCGCAGCGCCCTCTC-AGCAGAGAAAGGGGTGTGCTCAGT 658  
DB 863 GGCACATATTTGATGTGCGCAGCGTCTCCACCAAGTGGCAGAGGGGTGTGCTCAGT 922

QY 659 CATCAAGCTGAATAACA---GGCAGTTATTCAACGACGACAAAC--TATGATGG 713  
DB 923 CATCAAGCTGAATAACAAGGCAATATTATTCATATGTCATGCAAAACCTATAGTGG 982  
QY 714 ACATTT--AACAGTCTCCCGAGCAAGAACTATTCTATGCAATCAAGTTTACCAGGCT 771  
DB 983 ACATTTTAAACAGTCTCCCGAGCAAGAACTATTCTATGCAATCAAGTTTACCAGGCT 1042  
QY 772 CAGTGGCTTCAATATGAGATCTTTTGCATTTGATCAAGTGAAGAACTCACTGACAGAA 831  
DB 1043 CAGTGGCTTCAATATGAGATCTTTTGCATTTGATCAAGTGAAGAACTTACAGAGAG 1102  
QY 832 GAATTTATCTAGCTATGACCTAATTTGATGTTGATGTTGATGTTGATGTTGATGTTGAT 891  
DB 1103 GAATTTATCTAGCTATGACCTAATTTGATGTTGATGTTGATGTTGATGTTGATGTTGAT 1162  
QY 892 GTCTGCTCTCCAGATATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 951  
DB 1163 GTCTGCTCTCCAGATATGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1222  
QY 952 TCGCTCATAGCTCTTCTCTGATGATCAGAGGCTGCTGAGAGCCGTGTCAGAGAT 1011  
DB 1223 TCGCTCATAGCTCTTCTCTGATGATCAGAGGCTGCTGAGAGCCGTGTCAGAGAT 1282  
QY 1012 GAGCAGAGC---CAGAGAAACCTGCTGTGACATTTGAAATTAAGAACCGGAGAAC 1068  
DB 1283 GAAACAAACAAATTAAGAAAGAAATTAACCTGTAACCTTTGAATTAAGAACCGGAGAAC 1342  
QY 1069 TTGAGCGAGAGCTGTGAGCTGTGAGAACCGCCGACGCTCTTGGAGCAGCAGCG 1128  
DB 1343 TTGAGCGAGAGCTGTGAGCTGTGAGAACCGCCGACGCTCTTGGAGCAGCAGCG 1402  
QY 1129 AAGAGAGAGAGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1188  
DB 1403 AAGAGAGAGAGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1462  
QY 1189 CGCAGAGAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1248  
DB 1463 CGCAGAGAGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1522  
QY 1249 GAGCTGAGCGGAGCGAG 1308  
DB 1523 GAGCTGAGCGGAGCGAG 1582  
QY 1309 AAGAGAGAGAGCT-GGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1367  
DB 1583 AAGAGAGAGAGCTGAG 1642  
QY 1368 CTTGAATCAGAGAGAAACAG 1427  
DB 1643 ACTTAATCAGAGAGAAACAG 1702  
QY 1428 TCTGAGATTTGAGTTAGAGCTTGAATGCAAAAGCATCAGCTAGAGAGAGAGAGAGAG 1487  
DB 1703 TTTGGAATTTGAGTTAGAGCTTGAATGCAAAAGCATCAGCTAGAGAGAGAGAGAGAG 1762  
QY 1488 GATATCAGAGTGTGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1547  
DB 1763 AGATATCAGAGTGTGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1822  
QY 1548 AGAGCTAAGAAATTTGCTG--AAATCACCACCTTACAGAGAGAGT--GCAGAAATCTCAG 1603  
DB 1823 AGAGCTAAGAAATTTGCTG--AAATCACCACCTTACAGAGAGAGT--GCAGAAATCTCAG 1882  
QY 1604 AAA--TGCTTGAAGAGCTTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1660  
DB 1883 AAAATGCTTTGAAGAGCTTATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1942  
QY 1661 TCCAGAGAGAGAGCTTTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1719  
DB 1943 TCCAGAGAGAGAGCTTTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2002



QY 1720 AAGAGCTGGCCCCGAGCAGCTCCGGGAGCAGCTGAGCGAGGTGAGAGAGACCAGG 1779  
 Db 2003 AAGAGCTGGCCCCGAGCAGCTCCGGGAGCAGCTGAGCGAGGTGAGAGAGACCAGG 2062  
 QY 1780 TCAGAGCTGAGAGAGTGTATGTTTTCACACACAGCTGAGAGAGTGTGAGAGATCAT 1839  
 Db 2063 TCAGAGCTGAGAGAGTGTATGTTTTCACACACAGCTGAGAGAGTGTGAGAGATCAT 2122  
 QY 1840 AGCAAGCAGCAATCCAGAGCAGAGGTCCCTGAGGCGAGCGAGCTGAGAGCAGAAAGAG 1899  
 Db 2123 AATAGAGCAATCTCCAGAGCAGAAAGTCCATGAGAGCTGAGAGCTGAGAAACAGAAAGAA 2182  
 QY 1900 CAGAGAGAGAGAGCTGAGTGTAGAGAGCAAAAGAGAGCCTCAGAGAGAGATTGAG 1959  
 Db 2183 CAGAGAGAGAGAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAG 2242  
 QY 1960 GAAAGGAGCAGCAATGGCTGAGAGCATGTGACAGAGAG---GAGAGCGACGCCCCCG 2016  
 Db 2243 GAAAGGAGCAGCAATGGCTGAGAGCATGTGACAGAGAGAGAGAGAGAGAGAGAGAGAG 2302  
 QY 2017 AAG 2075  
 Db 2303 AAG 2362  
 QY 2076 AG 2135  
 Db 2363 GAG 2422  
 QY 2136 GCGAGTAACTGGAGCAG 2195  
 Db 2423 ACCAGCTAACTGGAGCAG 2482  
 QY 2196 TTCTGACAGAGAGAGTGTAAAGTGTATATATACAGAGAGAGAGAGAGAGAGAGAGAG 2255  
 Db 2483 TTCTGACAGAGAGAGTGTAAAGTGTATATATACAGAGAGAGAGAGAGAGAGAGAGAG 2542  
 QY 2256 AAGTCAAGTGAAGTTCACATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2300  
 Db 2543 AAGTCAAGTGAAGTTCACATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2602  
 QY 2301 GATGAAAGCAG 2360  
 Db 2603 GATGAAAGCAG 2662  
 QY 2361 ATGAGTCCCTGCAAACTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2420  
 Db 2663 GTGAGTCCCTGCAAACTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2722  
 QY 2421 ACCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 2480  
 Db 2723 ACCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 2782  
 QY 2481 TCCTTTGACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 2540  
 Db 2783 TCCTTTGACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 2842  
 QY 2541 TTCAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 2600  
 Db 2843 TTCAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 2902  
 QY 2601 GGCTCAG-CCCTTCTGAGCCGCTGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2659  
 Db 2903 AGCCAGAGCCCTTCTGAGCCGCTGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2962  
 QY 2660 CCGCAGAGCAG 2719  
 Db 2963 CCGCAGAGCAG 3022  
 QY 2720 AAGGAGTGAAG 2778  
 Db 3023 AAGGAGTGAAG 3082  
 QY 2779 TTTAACAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2838

Db 3083 TTTAACAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 3142  
 QY 2839 GTTCAAGTGAAGAGAGTGTGTCCCAAGTCTTACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 2898  
 Db 3143 GTTCAAGTGAAGAGAGTGTGTCCCAAGTCTTACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 3202  
 QY 2899 AGAAGTGAAGAGAGTGTGTCCCAAGTCTTACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 2957  
 Db 3203 AGAAGTGAAGAGAGTGTGTCCCAAGTCTTACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 3262  
 QY 2958 GGCTTCCCGGCGCCAG 3016  
 Db 3263 AGCTTCCCGGCGCCAG 3322  
 QY 3017 ACAGAGTGTGAG-CAAGAGATTTAACCTTT--CAGAGAGAGAGTGTGATTTGAGTT 3072  
 Db 3323 ACAGAGTGTGAGCAAGGAGATTTAACCTTTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3382  
 QY 3073 ACCAAGAAAGTGTGAGTGTGAG 3131  
 Db 3383 ACCAAGAAAGTGTGAGTGTGAG 3442  
 QY 3132 TTCTACTATGTGAGGCTTAAAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3191  
 Db 3443 TTCTACTATGTGAGGCTTAAAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3502  
 QY 3192 TTTAG 3250  
 Db 3503 TTTAG 3562  
 QY 3251 AAGCACTACCTGAGCTGTGAG 3309  
 Db 3563 AGCACTACTCTGAGCTGTGAG 3622  
 QY 3310 TGTGAG 3369  
 Db 3623 TGTGAG 3682  
 QY 3370 AATTATGTCAGAGTGTGAG 3429  
 Db 3683 AATTATGTCAGAGTGTGAG 3742  
 QY 3430 ACCGAGTGAAG 3489  
 Db 3743 TCAAGAGATTTAG 3802  
 QY 3490 GATGAGAGTGAAGTGTGAG 3548  
 Db 3803 GATGAGAGTGAAGTGTGAG 3862  
 QY 3549 GAGTGTGAG 3608  
 Db 3863 TGAAGTGTGAG 3922  
 QY 3609 GCTGACACAGAGAT-GGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3641  
 Db 3923 GCTGACACAGAGATGTGAG 3956

## RESULT 8

AAZ34570 standard; cDNA; 5199 BP.

AAZ34570;

01-FEB-2000 (first entry)

Human SH3D1A cDNA.

DE SH3D1A gene; human; Down's syndrome; leukemia; cancer;  
 XX megakaryocytic abnormality; myeloproliferative disorder;  
 KW platelet disorder; neutral disorder; thrombocytopenia;  
 KW





DB 3415 ACAGACCCACCTTAAGTCAACAGCATTAGCGCGAGTGTGCGAGTGTGGATGTACGAC 3474  
QY 3475 TACACCGCCAGCAAGCATGACGAACTAGCCTTCAGCAAGGCGAGATCATCAAGCTCC 3534  
DB 3475 TACACCGCCAGCAAGCATGACGAACTAGCCTTCAGCAAGGCGAGATCATCAAGCTCC 3534  
QY 3535 AACAGAGAGACCCGAGCTGTGTGAAGAAGAGTCAAGTGGGCAAGTTGGCTTTCCCA 3594  
DB 3535 AACAGAGAGACCCGAGCTGTGTGAAGAAGAGTCAAGTGGGCAAGTTGGCTTTCCCA 3594  
QY 3595 TCCAAATTATGTAAAGTGCACACAGCATGACCCGACGACGAATGA 3642  
DB 3595 TCCAAATTATGTAAAGTGCACACAGCATGACCCGACGACGAATGA 3642  
RESULT 9  
ID ABK43498 standard; cDNA, 3319 BP.  
XX  
AC ABK43498;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE DNA encoding novel central nervous system protein #78.  
XX  
KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;  
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;  
KW cardiac arrest; cerebrovascular disorder; ischemia; angiogenesis;  
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;  
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;  
KW adenocarcinoma; reproductive system disorder; testicular feminisation;  
KW endocrine disorder; diabetes; cancer; leukemia; neovascularisation;  
KW respiratory disorder; renal disorder; kidney failure; blood disorder;  
KW myocardial infarction; wound healing; cell proliferation; skin aging;  
KW food additive; food preservative; gene therapy; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO20015318-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01332.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236337.  
PR 29-SEP-2000; 2000US-0236337.  
PR 29-SEP-2000; 2000US-0236337.  
PR 29-SEP-2000; 2000US-0236338.  
PR 29-SEP-2000; 2000US-0236339.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.

PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251889.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;

XX WPI, 2001-581633/65.

DR P-PSDB; AUB87168.

XX New isolated nucleic acid encoding a protein for diagnosing,  
 PT preventing, treating or ameliorating medical conditions and used as  
 food additives or preservatives -

XX Claim 1; SEQ ID No 88; 837pp; English.

XX The invention describes an isolated nucleic acid molecule (I) encoding a  
 CC novel central nervous system protein. (I) and polypeptides (III) encoded  
 CC by (I), are used to treat a medical conditions and in diagnosis of a  
 CC pathological condition. Disorders which are diagnosed or treated include  
 CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative  
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders  
 CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,  
 CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and  
 CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses  
 CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders  
 CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,  
 CC adenocarcinomas and irritable bowel syndrome, reproductive system  
 CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes  
 CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.  
 CC leukaemia, disorders involving neovascularisation e.g. malignancies,  
 CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.  
 CC acute kidney failure and blood related disorders e.g. myocardial  
 CC infarction. The polypeptides can also be used to aid wound healing and  
 CC epithelial cell proliferation, to prevent skin aging due to sunburn, to  
 CC maintain organs before transplantation, for supporting cell culture of  
 CC primary tissues, to regenerate tissues and in chemotaxis. The  
 CC polypeptides can also be used as a food additive or preservative to  
 CC increase or decrease storage capabilities, fat content, lipid, protein,

Query Match 56.1%; Score 2043.8; DB 23; Length 3319;  
 Best Local Similarity 85.2%; Pred. No. 0;  
 Matches 2324; Conservative 0; Mismatches 387; Indels 18; Gaps 3;

QY 1 ATGGCTGAGTTCCACACCTTTCGGGTGAGCTGATCTCGGACCATACCTGAG 60  
 DB 159 ATGGCTGAGTTCCACACCTTTCGGGTGAGCTGATCTCGGACCATACCTGAG 218  
 QY 61 GAAAGGCCAAGCATGACGACGAGCTTCTTACCTGAGCCGATAGCGGATTTACT 120  
 DB 219 GAAAGGCCAAGCATGATGACGAGCTTCTTACCTGAGCCGATAGCTGATTTACT 278  
 QY 121 GGTATCAAGGAGGAACTTTTTCCTCAATCTGGGTACCTCAGCTCTTACGACAA 180  
 DB 279 GGTATCAAGGAGGAACTTTTTCCTCAATCTGGGTACCTCAGCTCTTACGACAA 338  
 QY 181 ATATGGGCTAGCGGACATGAAATACGATGAAAGATGATCAATGAAATTTCCATA 240  
 DB 339 ATATGGGCTAGCTGACATGAAATACGATGAAAGATGATCAATGAAATTTCCATA 398  
 QY 241 GCCATGAAGCTTATCAAACTGAGCTCAAGATATAGCTCCCTCCACACTTCCCTCC 300  
 DB 399 GCTATGAAACTTATCAAACTGAGCTCAAGATATAGCTCCCTCCACACTTCCCTCC 458  
 QY 301 GTCATGAAACAGCAACGAGTGGCTATTCCAGTGCACACGATTTGGTATAGAGGATT 350  
 DB 459 GTCATGAAACAGCAACGAGTGGCTATTCCAGTGCACACGATTTGGTATAGAGGATT 518  
 QY 351 GCTAGCATGCCACCACTGACGCTGTTGCTCTGTCGCAATGGGCTCATTCCAGTTGTT 420  
 DB 519 GCCAGCATGCCACCGCTTACGAGCTGTTGCTCAGTGCATGAGGATTCATTCCAGTTGTT 578  
 QY 421 GGAATGCTCCACCTTATGATCTTCTGTCCTCCAGACAGAGGCTCCCTGGCTTAAC 480  
 DB 579 GGAATGCTCCACCTTATGATCTTCTGTCCTCCAGACAGAGGCTCCCTGGCTTAAC 638  
 QY 481 GGGGCTCTCCCGTCATACAGCTCTGCTGCTGGTTCGATCCTCGACGACATGGGCCA 540  
 DB 639 GGGGCTCTCCCGTCATACAGCTCTGCTGCTGGTTCGATCCTCGACGACATGGGCCA 698  
 QY 541 AAGAGTCTCTCTTCCAGACGATCTGTCGAGGTCACAACTTAATACCTAGTTACGAG 600  
 DB 699 AAGAGTCTCTCTTCCAGACGATCTGTCGAGGTCACAACTTAATACCTAGTTACGAG 758  
 QY 601 GCACATCATTCGATGTCGCGCAGCGCCCTCCAGACAGAAATGAGCTGCTCAGTCA 660  
 DB 759 GCACATCATTCGATGTCGCGCAGCGCCCTCCAGACAGAAATGAGCTGCTCAGTCA 818  
 QY 661 TCAGAGCTGAATACAGGAGCTTATTCACAGCGACGACAAACTATGAGTGAACCTTA 720  
 DB 819 TCAGAGCTGAATACAGGAGCTTATTCACAGCGACGACAAACTATGAGTGAACCTTA 878  
 QY 721 AC-----AGGTCGCCAGCAAGAACTATTCTCATGTCATCAATGATTAACCCAG 768  
 DB 879 ACAGGTCCTGTTTAGTTCGCCAGCAAGAACTATTCTCATGTCATCAATGATTAACCCAG 928  
 QY 769 GCTCAGCTGAGCTTCAATATGGAATCTTTTTCGACATTGATCAATGATTAACCCAG 828  
 DB 929 GCTCAGCTGAGCTTCAATATGGAATCTTTTTCGACATTGATCAATGATTAACCCAG 988  
 QY 829 GAAAGATTATTCATGAGTACCTAATGATGATGTCATGTCAGTCCAGCACTCCG 888  
 DB 999 GAAAGATTATTCATGAGTACCTAATGATGATGTCATGTCAGTCCAGCACTCCG 1058  
 QY 889 CCGGCTCGCTCCAGAAATACATCCCTCTTCTTCGAAAGAGTTGCTCCGCGAGTGG 948  
 DB 1059 CCGGCTCGCTCCAGAAATACATCCCTCTTCTTCGAAAGAGTTGCTCCGCGAGTGG 1118  
 QY 949 ATGTCGCTATTAAGCTCTTCTTCTGATGATGAGGCTGCTGAGGCGCTGTCAG 1008  
 DB 1119 ATATGTCATTAAGCTCAACATCTGATGATGAGGCTGCTGAGGCGAGCACTTTTGA 1178  
 QY 1009 GATGAGCAGCAGC---CAGAGAAAGAACTGCTGTCATTTGAAGATTAAGACGGAG 1065

```

Db 1179 GATGAAACAACAATTTAAGAAAAGAAATTAACCTGTAACTTTGAAGATTAAGAGCGGAG 1238
Oy 1066 AACTTGCAGCGAGCGAGCTGTGAGAGTGAAGACCGCGCAAGCCCTCTTGAGCAGAG 1125
Db 1239 AACTTTGAACGTGGCAACTGTGGAATCTGGAGAAACGAAGGCAAGCTCTCTGGAACAGAG 1298
Oy 1126 CGCAAGAGCAGAGCGGTTGGCTCAGCTGAGCGCGCCGAGCAGAGAGAAGAAAGCGG 1185
Db 1299 CCAAGAGACAGAGAGCGCTGGCCAGCTGAGCGGGCGAGCAGAGAGAGAGAGAGCGT 1358
Oy 1186 GAGCGCCAGAGACAGAGAGCGCCAGCTGAGCTGGAAGAACAGCTGAGAGAGAG 1245
Db 1359 GAGCGCCAGAGACAGAGAGCGCAAAAGACAATCTGGAACCACTGGAAAAAGAG 1418
Oy 1246 CGGAGCTGAGCGGAGAGAGAGAGAGAGAGAGAGAGATGAGAGAGCGGAGCGC 1305
Db 1419 CGGAGCTGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCT 1478
Oy 1306 GCAAAAACGGGAACTGGAAAAGGAGCGACAATTGAAATGGGAACGGAAACGAGAGAA 1365
Db 1479 GCAAAAACGGGAACTGGAAAAGGAGCGACAATTGAGTGGGAACGGAAATCGAAGCAAGA 1538
Oy 1366 CTCCTGAATCAAGAGAAACAAGAGAGAGAGCGGACCGTGTCTTGAAGCGAAGAGAG 1425
Db 1539 CTACTAAATCAAAAGAAACAAAGAACAAAGAGCATAGTTGTACTGAAAAGCAAAAGAAAAG 1598
Oy 1426 ACTCTGGAATTTGAGTTAGAACTGTAAGCTGTAACAAAAGCATCAGCTGGAAGGAAACCT 1485
Db 1599 ACTTTGAAATTTGAATTTAAGCTCTAAATGATPAAAAGCATCAACTGAAGAGGAACTT 1658
Oy 1486 CAGATATCAGGTGTGCTGAGTGGCAACCCAGAGCGAAAGAAATTTAGAGAGACGAACAGTCT 1545
Db 1659 CAAAGATATCAGATGTGATTTGAACCAACCAAGGCAAGAAATTTAGAGAGCAAAACAAATCT 1718
Oy 1546 AAGAGCTAAGAAATTTGCTGAATCACCCACTTACAGCAGCAGTTGAGAGATCTCAGCAA 1605
Db 1719 AAGAGCTAAGAAATTTGCTGAATCACCCACTTACAGCAGCAGTTGAGAGATCTCAGCAA 1778
Oy 1606 ATGCTTGAAGACTTATTCAGAGAAACGATACCTCAGTACGACAGTTAAACAACTCCAG 1665
Db 1779 ATGCTTGAAGACTTATTCAGAGAAACGATACCTCAGTACGACAGTTAAACAACTCCAG 1838
Oy 1666 CAGAACAGTTTGCATAGAGACTCGCTTCTTACCTCAAAAGAGCCTTGGAAAGCAAGAGAG 1725
Db 1839 CAGAACAGTTTGCATAGAGACTCGCTTCTTACCTCAAAAGAGCCTTGGAAAGCAAGAGAG 1898
Oy 1726 CTGGCCCGGAGAGAGTCCGGAGAGAGCTGGAACGAGGTGGAGAGAGAGACCAAGTCAAG 1785
Db 1899 CTAGCTCGGAGAGAGTCCGGAGAGAGCTGGAACGAGGTGGAGAGAGAGACCAAGTCAAG 1958
Oy 1786 CTGGAGGAGATTTGATGTTTCAACAACCGCTGAAGGAACTGAGAGATACATAGCAAA 1845
Db 1959 CTACAGGAGATTTGATGTTTCAACAACCGCTGAAGGAACTGAGAGATACATAGCAAA 2018
Oy 1846 CAGCAACTCCAGAGCAGAGAGTCCCTGAGAGAGCGCGCATGAACAGAGAAAGCAGAGAG 1905
Db 2019 CAGCAACTCCAGAGCAGAGAGTCCCTGAGAGAGCGCGCATGAACAGAGAAAGCAGAGAG 2078
Oy 1906 AAGAGAGCGCTGAGTTAGAGAGCAAAAGAGAGAGCTCAGAGAGCGAGTTGAGAGAGAG 1965
Db 2079 CGAAAGATCATAGAAATTTAGAAAAACAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 2138
Oy 1966 GACAGGCAATGGCTGGAGATGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2022
Db 2139 GACAGGCAATGGCTGGAGATGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2198
Oy 2023 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2082
Db 2199 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2258
Oy 2083 GCCAAGCCGAAAATGCAAGACAGAGAGTGGCTTTTTCATCCGATCAGAGAGCGAGCT 2142

```

```

Db 2259 GGCAAACAGAGAGACAGAGCAAGAGCTGGGTGGCTTTTCCATCAACACAGAGACGCT 2318
Oy 2143 AAGCTGGCCACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2202
Db 2319 AAGCAGAGCTGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2378
Oy 2203 CAGAGAGAGTAAAGAGTGTATATTAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2262
Db 2379 CAGAGAGAGTAAAGAGTGTATATTAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2438
Oy 2263 GATGATACCAATCCAGAGAGAGAGATATGATATGATGATGATGATGATGATGATGATGATGATGAT 2322
Db 2439 GATGATACCAATCCAGAGAGAGAGATATGATATGATGATGATGATGATGATGATGATGATGATGAT 2498
Oy 2233 CAGAGATGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2382
Db 2499 CCGCGCTGGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2558
Oy 2283 GAAAAAGTTCCAGAGAAATGAGTTCCCACTCCAGCCAAACAGTGAACCGATCTGACATCT 2442
Db 2559 GAGAAATCCAGAGAAATGAGTTCCCGCTCCAGTGAACCAAGTGAATGATCAATCT 2618
Oy 2443 GCGCCCTGCCCCCAAACTGGCTGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2502
Db 2619 GCGCCCTGCCCCCAAACTGGCTGCGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2678
Oy 2503 GAGCCCTCCCAAAACCCCAACCACTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2562
Db 2679 GAGCCCTCCCAAAACCCCAACCACTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2738
Oy 2563 AACGAGAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2622
Db 2739 AATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2798
Oy 2623 CTTAGTGTCTGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2682
Db 2799 CCAAGTGTCTGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2858
Oy 2683 TCCCATCTCTCCGCTGGGCGCAGAGGTA 2711
Db 2859 TCCCATCTCTCTGTGTAGGCGAGGCTGA 2887

RESULT 10
AA163825
ID AA163825 standard; cDNA; 3466 BP.
XX
AC AA163825;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 33.
XX
KW Human; antiarthritic; antirheumatic; antiproliferative; vasotropic;
KW cerebrioprotective; nootropic; neuroprotective; antibacterial; virucide;
KW fungicide; ophthalmological; cytostatic; immunosuppressive; nootropic;
KW neuroprotective; anti-allergic; hepatotropic; antidiabetic;
KW anti-inflammatory; anti-ulcer; anticonvulsant; antibacterial;
KW antiparastic; cardiac; gene therapy; cancer; immune disorder;
KW cardiovascular disorder; neurological disease; infection; human; ss.
OS Homo sapiens.
XX
PN MO200155308-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001MO-US01309.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.

```





CC the encoded proteins (AAM43497-AAM43660) useful for preventing, treating  
 CC or ameliorating medical conditions e.g. by protein or gene therapy. The  
 CC genes were isolated from a range of human tissues disclosed in the  
 CC specification. The nucleic acids, proteins, antibodies and (ant)agonists  
 CC are useful in the diagnosis, treatment and prevention of: (a) cancer,  
 CC e.g. breast and ovarian cancer and other cancers of the adrenal gland,  
 CC bone, bone marrow, breast, gastrointestinal tract, liver, lung, or  
 CC urogenital; (b) immune disorders e.g. Addison's disease, allergies,  
 CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,  
 CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 3466 BP, 1056 A; 829 C; 861 G; 710 T; 10 other;

Query Match 56.1%; Score 2043.8; DB 22; Length 3466;

Best Local Similarity 85.2%; Pred. No. 0;

Matches 2224; Conservative 0; Mismatches 387; Indels 18; Gaps 3;

QY 1 ATGGCTCAAGTTCACACCTTTGCGTGTAGCTGTGATGCTGCGCCATTAAGCTGAG 60  
 DB 277 ATGGCTCAAGTTCACACCTTTGCGTGTAGCTGTGATGCTGCGCCATTAAGCTGAG 336  
 QY 61 GAAAGGGCCCAACATGACAGAGATTCCTTACCTGGAAGCCGATGCGGANTTTACT 120  
 DB 337 GAAAGGGCCCAACATGACAGATTCCTTACCTGGAAGCCGATGCGGANTTTACT 396  
 QY 121 GGTGATCAAGCAGAGAACTTTTTCATCTGGGTTAAGCTGAGCTGCTTACAGAA 180  
 DB 397 GGTGATCAAGCAGAGAACTTTTTCATCTGGGTTAAGCTGAGCTGCTTACAGAA 456  
 QY 181 ATATGGGCGCTTACCGGACATGAATAACATGAGAGATGATCAAGTGAATTTCCATA 240  
 DB 457 ATATGGGCGCTTACCGGACATGAATAACATGAGAGATGATCAAGTGAATTTCCATA 516  
 QY 241 GCCATGAACCTTATCAAACTGAAAGCTACAGATATACGCTCCCTCCACACTTCCCT 300  
 DB 517 GCCATGAACCTTATCAAACTGAAAGCTACAGATATACGCTCCCTCCACACTTCCCT 576  
 QY 301 GTCATGAACACAGCAACCACTGCTATTTCCAGTGCACAGCAATTTGATAGAGAGATT 360  
 DB 577 GTCATGAACACAGCAACCACTGCTATTTCCAGTGCACAGCAATTTGATAGAGAGATT 636  
 QY 361 GCTAGCATGCAACCACTGCTATTTCCAGTGCACAGCAATTTGATAGAGAGATT 420  
 DB 637 GCTAGCATGCAACCACTGCTATTTCCAGTGCACAGCAATTTGATAGAGAGATT 696  
 QY 421 GGAATGTCTCCACCTTATGATCTTCTGCTCCAGAGCAGTGTCTCCCTGCTTAAC 480  
 DB 697 GGAATGTCTCCACCTTATGATCTTCTGCTCCAGAGCAGTGTCTCCCTGCTTAAC 756  
 QY 481 GGGGCTCTCCGCTCATAGAGCTGCTGCTGCTTGGCCATCTCCAGCCACATGAGCA 540  
 DB 757 GGGGCTCTCCGCTCATAGAGCTGCTGCTGCTTGGCCATCTCCAGCCACATGAGCA 816  
 QY 541 AAGAGTCTTCTTCCAGAGATCTGCTCAGAGTCACTTAATACTTAAGTTACAGAA 600  
 DB 817 AAGAGTCTTCTTCCAGAGATCTGCTCAGAGTCACTTAATACTTAAGTTACAGAA 876  
 QY 601 GACATATCTTGGATGCTGCTCCAGAGCCCTCCAGAGCAGAAATGGCTGTGCTCAGTCA 660  
 DB 877 GACATATCTTGGATGCTGCTCCAGAGCCCTCCAGAGCAGAAATGGCTGTGCTCAGTCA 936  
 QY 661 TCAAGGCTCAATACAGAGAGTATTAACAGCAGCAAGAACTATGATGAGACATTA 720  
 DB 937 TCAAGGCTCAATACAGAGAGTATTAACAGCAGCAAGAACTATGATGAGACATTA 996  
 QY 721 AC-----AGGTCCCAAGGCAAGAACTATTTCTCATGCAATCAAGTTACCCAG 768

DB 997 ACAGGTTCTGTTAGCTCCCAAGAGAACTATTTATGAGTCAAGTTTACACAG 1056  
 QY 769 GCTCAGCTGCTTCAATATGAAATCTTTTTCGACATGATCAAGATGAAACATCTGCA 828  
 DB 1057 GCTCAGCTGCTTCAATATGAAATCTTTTTCGACATGATCAAGATGAAACATCTGCA 1116  
 QY 829 GAAGAAATTTATCTTACGATGACCTTAATGATGTTGCAATGCTGCTGACACCTGCC 888  
 DB 1117 GAGAAATTTATCTTGGCAATGACCTTATGATGTTGATGCTGCTGACACCTGCCA 1176  
 QY 889 CCGCTCTGCTCCAGAAATGACCTTCTTCCAGAAAGTGTGCTCCGCGAGTGGG 948  
 DB 1177 CCGCTCTGCTCCAGAAATGACCTTCTTCCAGAAAGTGTGCTCCGCGAGTGGG 1236  
 QY 949 ATGCTCTCAAGCTCTTCTTCTGATCAAGAGCTGCTGAGAGCGCTGCTGAGAG 1008  
 DB 1237 ATATCTGCTTAAGCTCAATCTGATGATGAGAGCTACCAAGAGAACCAAGTTTAGAA 1296  
 QY 1009 GATGAGCAGAGC--CAGAGAAAGAACTGCTGTGACATTTGAAATAGAAAGCGGAG 1065  
 DB 1297 GATGAAACAAACAAATTAGAAAGAAATTAAGCTGTAAAGCTTTGAAAGAAAGCGGAG 1356  
 QY 1066 AACTTTCAGAGAGCAGTGTGAGCTGAGAAAGCGCCGCAAGCGCTTTGAGCAGCAG 1125  
 DB 1357 AACTTTCAGAGAGCAGTGTGAGCTGAGAAAGCGCGCAAGCGCTTTGAGCAGCAG 1416  
 QY 1126 CGCAAGAGCAGAGCGGTTGGCTCAAGCTGAGAGCGCGCAGCAGAGAGAAAGACGG 1185  
 DB 1417 CGCAAGAGCAGAGCGGTTGGCTCAAGCTGAGAGCGCGCAGCAGAGAGAAAGACGG 1476  
 QY 1186 GAGCGCAGAGAGAGAGCGCAAGCGGCAAGCTGAGAGAGCACTGAGAAAGCAG 1245  
 DB 1477 GAGCGCAGAGAGAGAGCGCAAGCGGCAAGCTGAGAGAGCACTGAGAAAGCAG 1536  
 QY 1246 CGGAGCTGAGAGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1305  
 DB 1537 CGGAGCTGAGAGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1596  
 QY 1306 GCAAAAAGGAACTGAAAGAGCAGCAGCACTTGAATGGGAAAGGAAAGCGGAGACAGAA 1365  
 DB 1597 GCAAAAAGGAACTGAAAGAGCAGCAGCACTTGAATGGGAAAGGAAAGCGGAGACAGAA 1656  
 QY 1366 CTCCTGAATCAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1425  
 DB 1657 CTCCTGAATCAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1716  
 QY 1426 ACTCTGAGATTTGATGAGAGAGCTGGAATGACAAAAGAGATCAAGTGAAGAAACTT 1485  
 DB 1717 ACTCTGAGATTTGATGAGAGAGCTGGAATGATGAAGATCAAGTGAAGAAACTT 1776  
 QY 1486 CAGGATATCAGGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1545  
 DB 1777 CAGGATATCAGGATGCTGGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1836  
 QY 1546 AGAGAGTAAAGATTTGTAATGACCACTTACAGAGAGAGAGAGAGAGAGAGAGAG 1605  
 DB 1837 AGAGAGTAAAGATTTGTAATGACCACTTACAGAGAGAGAGAGAGAGAGAGAGAG 1896  
 QY 1606 ATGCTTGAAGATTTATTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1665  
 DB 1897 ATGCTTGAAGATTTATTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1956  
 QY 1666 CAGAAAGTTTGCATAGAGAGCTGCTTCTTAACCTCAAAAGAGCTTTGAGAAAGAGAG 1725  
 DB 1957 CAGAAAGTTTGCATAGAGAGCTGCTTCTTAACCTCAAAAGAGCTTTGAGAAAGAGAG 2016  
 QY 1726 CTGCGCCGAGAGAGCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1785  
 DB 2017 CTGAGCTCGGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2076  
 QY 1786 CTGAGAGAGATTTGATTTTCAACAAACAGCTGAGAGAGAGAGAGAGAGAGAGAGAG 1845

```

Db 2077 CTACAGAGATTGATTTTCAATATACAGCTGAGGAACTAAGAGAAATACCAATTAAG 2136
Qy 1846 CAGCACTCCAGAAAGCAGAGAGTCCCTGAGAGCAGCGGACTGAAGAGAAAGAGCAGAG 1905
Db 2137 CAACAATCCAGAAAGCAAAAGTCCATGAGGCTGAACGACTGAACAAGAAAGAAAGAA 2196
Qy 1906 AGGAAGAGCTGAGATTAGAGAAAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1965
Db 2197 CGAAGATCATAGATTAGAAAAAAGAAAAAGAAAGAAAGAAAGAGAGAGAGAGAG 2256
Qy 1966 GACAAGCAATGGCTGAGAGCATGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2022
Db 2257 GACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2316
Qy 2023 CACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2082
Db 2317 CACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2376
Qy 2083 GCCAAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2142
Db 2377 GCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2436
Qy 2143 AAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2202
Db 2437 AAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2496
Qy 2203 CAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2262
Db 2497 CAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2556
Qy 2263 GATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2322
Db 2557 GATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2616
Qy 2323 CAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2382
Db 2617 CCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2676
Qy 2383 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2442
Db 2677 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2736
Qy 2443 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2502
Db 2737 GCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2796
Qy 2503 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2562
Db 2797 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2856
Qy 2563 AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2622
Db 2857 AATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2916
Qy 2823 CCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2882
Db 2917 CCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2976
Qy 2683 TCCCATCTCCGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2711
Db 2977 TCCCATCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3005

```

```

RESULT 11
AAZ34573
ID AAZ34573 standard; cDNA; 2079 BP.
XX
AC AAZ34573;
XX
DT 01-FEB-2000 (first entry)
XX
DE Human SH3D1A cDNA clone 5.
XX

```

```

KW SH3D1A gene; human; Down's syndrome; leukemia; cancer;
KW megakaryocytic abnormality; myeloproliferative disorder;
KW platelet disorder; neural disorder; thrombocytopenia;
KW haematopoietic disorder; cognitive dysfunction; microcephaly;
KW lisencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy;
KW ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 136..2079
FT FT /*tag= a
FN W09953062-A2.
PD 21-OCT-1999.
PX 16-Apr-1999; 99WO-US08371.
PR 16-Apr-1998; 98US-0082007.
PA (CEDA-) CEDARS SINAI HEALTH SYSTEM.
PI Korenberg JR. Chen X;
PX WPI; 1999-633829/54.
DR P-PSDB; AAY32157.
PT Nucleic acid from the human SH3D1A gene and its products, useful for
PT the diagnosis and treatment of myeloproliferative disorders and
PT leukemia -
PS Claim 2, Fig 12, 99PP, English.
XX
XX This is the nucleotide sequence of a cDNA clone, termed clone 5,
XX corresponding to a novel human SH3 gene, termed the SH3D1A gene,
XX that contributes to the development of platelets and the
XX pathogenesis of leukemias, both in general and in particular those
XX involving the megakaryocytic lineage. The SH3D1A gene maps to the
XX small candidate region for low platelets on chromosome 21.
XX Sequencing of 5 different sizes of cDNA clone from foetal brain
XX (see AAZ34570-74) suggests that at least 3 isoforms exist. The
XX invention provides methods for the diagnosis and treatment of
XX megakaryocytic abnormality, myeloproliferative disorder, platelet
XX disorder, acute leukemia, neural disorders, thrombocytopenia,
XX platelet disorder on chromosome 21, low platelets in deletion for
XX 21, association of gains in chromosome 21 with leukemias, neural
XX abnormalities, dysfunctions and disorders including brain
XX malformations and corresponding cognitive dysfunctions.
XX microcephaly, lisencephaly, and colpocephaly. Methods are also
XX provided for: suppressing cells unable to regulate themselves;
XX screening for a somatic alteration in the SH3D1A gene; monitoring
XX the progress and adequacy of a treatment; monitoring tumour risk
XX progress or megakaryocytic abnormality; myeloproliferative disorder,
XX haematopoietic disorder, platelet disorder or leukemia; and
XX treatment of a subject (including a prenatal subject) having
XX megakaryocytic abnormality, myeloproliferative disorder,
XX platelet disorder, leukemia or neural disorder using a
XX nucleic acid that expresses SH3D1A or its antisense nucleic acid.
XX
SQ Sequence 2079 BP; 700 A; 451 C; 495 G; 433 T; 0 other;

```

```

Query Match 41.0%; Score 1495; DB 20; Length 2079;
Best Local Similarity 85.9%; Pred. No. 0;
Matches 1670; Conservative 0; Mismatches 270; Indels 3; Gaps 1;
Qy 1 ATGGCTAGTTTCCCAACCTTTGGTGTGAGCTGAGATGCTGAGGCAATTAAGTGTGAG 60
Db 136 ATGGCTAGTTTCCCAACCTTTGGTGTGAGCTGAGATGCTGAGGCAATTAAGTGTGAG 195
Qy 61 GAAAGGCAAGCATGACAGCAGTCTTGAAGCGATGAGGAGATTTATTACT 120
Db 196 GAAAGGCAAGCATGACAGCAGTCTTGAAGCGATGAGGAGATTTATTACT 255

```

QY 121 GGTGATCAAGCGAGAACTTTTTTTCATCTGGTTACCTGAGCTGTCTTAGACA 180  
 DB 256 GGGATCAAGCTAGAAACTTTTTTTCATCTGGTTACCTCAACTGTTTAGACAG 315  
 QY 181 ATATGGCGCTGCGGACATGAAATAGATGGAAGATGATCAATGGAATTTTCCATA 240  
 DB 316 ATATGGGCACTAGCTACATGAAATAGATGGAAGATGATCAATGGAATTTTCCATA 375  
 QY 241 GCCATGAGCTTATCAAACTGAAAGTACAAAGATACAGTCCCTCCACATTCCTCCCT 300  
 DB 376 GGTATGAACCTTATCAAACTGAAAGTACAAAGATACAGTCCCTCCACATTCCTCCCT 435  
 QY 301 GTCATGAAACAGCAACCAAGTGGCTATTTCCAGTGCACGACATTTGGTATGAGAGGAT 360  
 DB 436 GTCATGAAACAGCAACCAAGTGGCTATTTCCAGTGCACGACATTTGGTATGAGAGGAT 495  
 QY 361 GGTAGATGCGCAACTGACAGTGTGTCTCTGTGCGCAATGGGCTCCATTCAGATTTGT 420  
 DB 496 GCGAGATGCGCAACTGACAGTGTGTCTCTGTGCGCAATGGGATCCATTCAGATTTGT 555  
 QY 421 GGAATGTCTCACCCCTTATCTTCTGTCTCTCCAGCAGAGTGGCTCCCTGGCTAAC 480  
 DB 556 GGAATGTCTCACCCCTTATCTTCTGTCTCTCCAGCAGAGTGGCTCCCTGGCTAAC 615  
 QY 481 GGGGCTCTCCGTCATACAGGCTCTGCTGGTGGTGGCATCTTGACGACATGGCCA 540  
 DB 616 GGGGCTCTCCGTCATACAGGCTCTGCTGGTGGTGGCATCTTGACGACATGGCCA 675  
 QY 541 AAGAGTCTCTCTTCAGCAGATCTGTCCAGGCTCAAACTTAAACACTAAGTTACAGAAG 600  
 DB 676 AAGAGTCTCTCTTCAGCAGATCTGTCCAGGCTCAAACTTAAACACTAAGTTACAGAAG 735  
 QY 601 GCACATCATCTGATGTCCGCAAGCGCCCTCCAGCAGAGTGGGCTGTGCTCAGTCA 660  
 DB 736 GCACATCATCTGATGTCCGCAAGCGCCCTCCAGCAGAGTGGGCTGTGCTCAGTCA 795  
 QY 661 TCAAGGCTGAAATACAGGAGTTATTCACAGCCAGCAAAAATAATGATGAGACATTA 720  
 DB 796 TCAAGGCTGAAATACAGGAGTTATTCACAGCCAGCAAAAATAATGATGAGACATTA 855  
 QY 721 ACAGGCTCCCGAGCAAGAACTATTCATGCAATCAAGTTTACCCAGGCTCAGCTGCT 780  
 DB 856 ACAGGCTCCCGAGCAAGAACTATTCATGCAATCAAGTTTACCCAGGCTCAGCTGCT 915  
 QY 781 TCAATATGGAATCTTCTGCAATGATCAAGATGAGAACTCAGCAGAGAAATTTATTC 840  
 DB 916 TCAATATGGAATCTTCTGCAATGATCAAGATGAGAACTTACAGCAGAGAAATTTATTC 975  
 QY 841 CTAGCTATGCACTTATGATGATGTCATGTCGTGTCAGCCACTGCGCGCTGCTGCT 900  
 DB 976 CTAGCTATGCACTTATGATGATGTCATGTCGTGTCAGCCACTGCGCGCTGCTGCT 1035  
 QY 901 CCAAGATACATCCCTCTCTCTCTCAGAGAGTTCCGCTCCGCAAGTGGATGTCCTCAT 960  
 DB 1036 CCAAGATACATCCCTCTCTCTCTCAGAGAGTTCCGCTCCGCAAGTGGATGTCCTCAT 1095  
 QY 961 AGCTCTTCTCTGAGATCAGAGGCTGCTGAGAGCGCTGCTCAGAGAGATAGAGCAG 1020  
 DB 1096 AGCTCTTCTCTGAGATCAGAGGCTGCTGAGAGCGCTGCTCAGAGAGATAGAGCAG 1155  
 QY 1021 C--CAGAGAGAACTGCTGTCATTTGAAGATTAAGAACCGGAGAACTTCAGCGA 1077  
 DB 1156 CAATTAGAAAAGAAATTTACTGTAACGTTTGAAGATTAAGAACCGGAGAACTTTGAACG 1215  
 QY 1078 GGCAGTGTGAGCTGAGAGAGCGCGCAAGGCTCTTGAGACAGCAGCGCAAGAGCAG 1137  
 DB 1216 GGCAGTGTGAGCTGAGAGAGCGCGCAAGGCTCTTGAGAGCAGCGCAAGAGCAG 1275  
 QY 1138 GAGCGGTTGGCTCAGCTGAGAGCGCGCGAGCAGAGAGAAAGCGGAGCGCAGAG 1197  
 DB 1276 GAGCGGTTGGCTCAGCTGAGAGCGCGCGAGCAGAGAGAAAGCGGAGCGCAGAG 1335

QY 1198 CAGAGGCGCAAGCGGCACTTGAGACTGAGAGACAGCTGAGAGAACGCGGAGCTGAG 1257  
 DB 1336 CAAGGCGCAAAAGAACTTGAGACTGAGAGAACCTGAGAGAACGCGGAGCTGAG 1395  
 QY 1258 CGGAGCGAGAGAGAGAGAGAGAGAGAGAGATGAGAGAGCGCGGAGCGCAAGCGGAA 1317  
 DB 1396 CGGAGCGAGAGAGAGAGAGAGAGAGAGAGAGATGAGAGAGCGCGGAGCGCAAGCGGAA 1455  
 QY 1318 CTGAAAAGCGCAACCACTTGAAATGGAAACGAAACCGAGACAGAACTCTGAATCAG 1377  
 DB 1456 CTGAAAAGCGCAACCACTTGAAATGGAAACGAAACCGAGACAGAACTCTGAATCAG 1515  
 QY 1378 AGGAACAAAG 1437  
 DB 1516 AGGAACAAAG 1575  
 QY 1438 GAGTTAGAGCTGAGATGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1497  
 DB 1576 GAGTTAGAGCTGAGATGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1635  
 QY 1498 TGTGAGCTGCGCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1557  
 DB 1636 TGTGAGCTGCGCAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1695  
 QY 1558 ATTGCTGAATCACCACTTACAGCAGCAGTGGAGAGATCTTACGAAATGCTTGAAGA 1617  
 DB 1696 ATTGCTGAATCACCACTTACAGCAGCAGTGGAGAGATCTTACGAAATGCTTGAAGA 1755  
 QY 1618 CTTATTCAGAGAAACGATACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 1677  
 DB 1756 CTTATTCAGAGAAACGATACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 1815  
 QY 1678 CATGAGACTGCTTCTTACCTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1737  
 DB 1816 CATGAGACTGCTTCTTACCTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1875  
 QY 1738 CAGCTCCGGAG 1797  
 DB 1876 CAGCTCCGGAG 1935  
 QY 1798 GATGTTTCAACCAACGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1857  
 DB 1936 GATGTTTCAACCAACGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1995  
 QY 1858 AAGCAGAGAGTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1917  
 DB 1996 AAGCAGAGAGTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2055  
 QY 1918 GAGTTAG 1940  
 DB 2056 GAGTTAG 2078  
 RESULT 12  
 AAK94139  
 ID AAK94139 standard; cDNA; 2131 BP.  
 XX  
 AC AAK94139;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human full-length cDNA, SEQ ID NO: 2646.  
 XX  
 KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EPI130094-A2.  
 XX  
 PD 05-SEP-2001.  
 XX  
 PF 07-JUL-2000; 2000BP-0114089.  
 XX

PR 08-JUL-1999: 99QP-0194486.  
PR 11-JAN-2000: 2000JP-0118774.  
PR 02-MAY-2000: 2000UP-0183765.  
XX  
XX (HELI-) HELIX RES INST.  
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y,  
PI Makamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;  
DR WPI, 2001-524255/58.  
XX P-PSDB; AAM93229.  
PT 830 Primers useful for synthesizing full length cDNA clones and their  
XX use in genetic manipulation -  
XX  
XX Claim 8; SEQ ID NO 2646; 1380bp + sequence listing; English.  
XX  
XX The invention relates to primers for synthesizing full length cDNA  
CC clones. 830 cDNA molecules encoding a human protein have been  
CC isolated and nucleotide sequences of 5' - and 3' - ends of the cDNA  
CC molecules have been determined. Primers for synthesizing the full length  
CC cDNA are useful for clarifying the function of the protein encoded by  
CC the cDNA. The full length clones were obtained by construction of full  
CC length enriched cDNA libraries that were synthesized by the oligo-capping  
CC method. The primers enable the production of the full length cDNA easily  
CC without any special methods. The present sequence is a full length  
CC human cDNA of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in CD-ROM format directly from EPO.  
XX  
XX Sequence 2131 BP; 694 A; 469 C; 526 G; 442 T; 0 other;  
SQ  
Query Match 40.7%; Score 1483.8; DB 22; Length 2131;  
Best Local Similarity 86.0%; Pred. No. 0;  
Matches 1657; Conservative 0; Mismatches 267; Indels 3; Gaps 1;  
QY 1 ATGGCTAGTTCCTCCACACCTTTGGGTGTAAGCTGATGCTGGGCGCAATCTGGAG 60  
DB 205 ATGGCTAGTTCCTCCACACCTTTGGGTGTAAGCTGATGCTGGGCGCAATCTGGAG 264  
QY 61 GAAAGGCGCCAGCATGACGACGATCTTGAAGCTGAGGCGGATGAGGATTTACT 120  
DB 265 GAAAGGCGCCAGCATGACGACGATCTTGAAGCTGAGGCGGATGAGGATTTACT 324  
QY 121 GGTGATCAAGCAGGAACTTTTTCATCTGGGTTACTTCAAGCTGCTTACGACAA 180  
DB 325 GGTGATCAAGCAGGAACTTTTTCATCTGGGTTACTTCAAGCTGCTTACGACAA 384  
QY 181 ATATGGGCGGCTAGGAGCATGAATPAAGATGAAGATGATGATCAAGTGAATTTCCATA 240  
DB 385 ATATGGGCGGCTAGGAGCATGAATPAAGATGAAGATGATGATCAAGTGAATTTCCATA 444  
QY 241 GCCATGAAGCTTATCAACTGAAGCTACAAAGATATCAGCTCCCTCCACACTTCCCT 300  
DB 445 GCTATGAACCTTATCAACTGAAGCTACAAAGATATCAGCTCCCTCCACACTTCCCT 504  
QY 301 GTCATGAACAGCAACGAGTGTATTTCCAGTGCACAGATTTGGTATAGAGGAGATT 360  
DB 505 GTCATGAACAGCAACGAGTGTATTTCCAGTGCACAGATTTGGTATAGAGGAGATT 564  
QY 361 GCTATGAAGCAGCACTCAGAGCTGTGCTCTGTCCTGCAATGGGCTCCATTCAGTTGT 420  
DB 565 GCTATGAAGCAGCACTCAGAGCTGTGCTCTGTCCTGCAATGGGCTCCATTCAGTTGT 624  
QY 421 GGAATGCTCCACCTTATGATCTTCTGCTCCCTCAGAGAGAGTCTCCCTGGCTAAC 480  
DB 625 GGAATGCTCCACCTTATGATCTTCTGCTCCCTCAGAGAGAGTCTCCCTGGCTAAC 684  
QY 481 GGGGCTCTCCGCTATCAGAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
DB 685 GGGGCTCTCCGCTATCAGAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 744  
QY 541 AAGAGTTCTTCTTCCAGAGATCTGTCCAGGCTCAGCAATTTAACTAAGTTACAGAG 600

DB 745 AAGAGTTCTTCTTCCAGAGATCTGTCCAGGCTCAGCAATTTAACTAAGTTACAGAG 804  
QY 601 GCAATCATTCGATGTCGCGAGCGCCCTCCAGCAGCAATGGGCTGCGCTGCTCA 660  
DB 805 GCAAGTATTTGATGTCGCGAGCGCTCCAGCAGCAATGGGCTGCTGCTGCTCA 864  
QY 661 TCAAGGCTGAATATCAGAGCAGTATTTCAAGCAGCAGCAAACTATGATGACACTTA 720  
DB 865 TCAAGGCTGAATATCAGAGCAGTATTTCAAGCAGCAGCAAACTATGATGACACTTA 924  
QY 721 AAGGCTCCCGCAGCAAGAACTATTTCTGATGATCAAGTTTACCCAGGCTGCTGCT 780  
DB 925 AAGGCTCCCGCAGCAAGAACTATTTCTGATGATCAAGTTTACCCAGGCTGCTGCT 984  
QY 781 TCAATATGATCTTTCTGATGATCAAGTTTACCCAGGCTGCTGCTGCTGCTGCT 840  
DB 985 TCAATATGATCTTTCTGATGATCAAGTTTACCCAGGCTGCTGCTGCTGCTGCT 1044  
QY 841 TTAAGTATGACCTTAATGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGAT 900  
DB 1045 CTGGCAATGACCTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1104  
QY 901 CCAAGATACATCCCTCTCTCTGAGAAAGTTGCTCCGCGAGTGGAGTGTCTGAT 960  
DB 1105 CCAAGATACATCCCTCTCTCTGAGAAAGTTGCTCCGCGAGTGGAGTGTCTGAT 1164  
QY 961 AGCTCTTCTGTCGATGATCAGAGGCTCTGAGAGGCTGTCGAGAGTGAAGTGAAGCAG 1020  
DB 1165 AGCTCAATCTGATGATGATCAGAGGCTCTGAGAGGCTGTCGAGAGTGAAGTGAAGCAG 1224  
QY 1021 C---CAGAGAAAGAACTGCTGTCGATGATGATGATGATGATGATGATGATGATGAT 1077  
DB 1225 CAATTGAAAGAAATTAATCTGTAACGTTGTAATGAAAGCGGGAATCTTTGAACT 1284  
QY 1078 GCGAGTGTGAGCTGAGAAAGCGCGCAAGCGCTTGGAGCAGCAGCAGCAAGAGCAG 1137  
DB 1285 GCGAGTGTGAGCTGAGAAAGCGCGCAAGCGCTTGGAGCAGCAGCAGCAAGAGCAG 1344  
QY 1138 GAGCGGTTGCTCAGCTGAGAGCGCGCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAG 1197  
DB 1345 GAGCGGTTGCTCAGCTGAGAGCGCGCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAG 1404  
QY 1198 CAGAGGCGCAAGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1257  
DB 1405 CAGAGGCGCAAGCGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1464  
QY 1258 CCGCAGCAG 1317  
DB 1465 CCGCAGCAG 1524  
QY 1318 CTGGAAGAGCAGCAGCACTTGAATGGGAACGGAACGGAACGGAACGGAACGGAACG 1377  
DB 1525 CTGGAAGAGCAGCAGCACTTGAATGGGAACGGAACGGAACGGAACGGAACGGAACG 1584  
QY 1378 AGGAACAAG 1437  
DB 1585 AGGAACAAG 1644  
QY 1438 GAGTTAAGAGCTCTGAATGACAAAAGCATGACTAGAGAGAGAGAGAGAGAGAGAGAG 1497  
DB 1645 GAGTTAAGAGCTCTGAATGACAAAAGCATGACTAGAGAGAGAGAGAGAGAGAGAGAG 1704  
QY 1498 TGTGAGTGGCAACCCAGAGCAGAGAAATTTGAGAGCAGCAAGTCTAGAGAGTAA 1557  
DB 1705 TGTGAGTGGCAACCCAGAGCAGAGAAATTTGAGAGCAGCAAGTCTAGAGAGTAA 1764  
QY 1558 ATTGCTGAATATCAGCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1617  
DB 1765 ATTGCTGAATATCAGCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1824  
QY 1618 CTTATTTCCAGAGAAACATATCTCAGTGAACAGTTTAAAGAGTCCAGCAGAGAGTTTG 1677

DB 1825 CTTATTCCAGAAAAACAGATCTCAATGACCAATTAAAAAAAGTTCCAGCAGAACGTTTG 1884  
QY 1678 CATAGAGACTCGCTTCTTACCCCTCAAAAGAGCCTTGAAGCAAGAGCTGGCCGCGAG 1737  
DB 1885 CACAGAGATTCACTTGTACCTTAACAAAGCCTTAGAAGCAAAAGAACTAGCTGGGAG 1944  
QY 1738 CAGCTCCGGAGAGAGCTGGACGAGGTGAGAGAGAGACCAAGCTCAAGTGCAGAGATT 1797  
DB 1945 CACCTTACGAGACCACTGATGATGAGAGAAAGAACTAGATCAAAACTACAGAGATT 2004  
QY 1798 GATGTTTCAACACAGCTGAAGAGAACTGAGAGATCATAGCAAAACAGACCTCCAG 1857  
DB 2005 GATATTTTCAATATATACCTGAGAGAACTTGAAGAAATACAAATAGCAACACTCCAG 2064  
QY 1858 AAGCAGAGTCCCTGGAGGACGAGCTGAAAGCAAAAGAGCAGAGAGAGAGCTTG 1917  
DB 2065 AAGCAAAAGTCCATGAGGCTGAGAGACTGAAACAGAAAGAAAGAAAGATCATTA 2124  
QY 1918 GAGTTAG 1924  
DB 2125 GAATTAG 2131

RESULT 13  
AAS84762  
ID AAS84762 standard; cDNA; 2874 BP.  
AC AAS84762;  
XX 13-FEB-2002 (first entry)  
DE DNA encoding novel human diagnostic protein #20566.  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.  
OS Homo sapiens.  
PN WO200175067-A2.  
XX 11-OCT-2001.  
PF 30-MAR-2001; 2001WO-US08631.  
XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
PI Drmanac RT, Liu C, Tang YT;  
XX MPI: 2001-639362/73.  
DR P-PSDB; ABG20575.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostic, forensic, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
PS Claim 1; SEQ ID No 20566; 103bp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2874 BP; 940 A; 679 C; 703 G; 552 T; 0 other;  
Query Match 35.5%; Score 1291.4; DB 23; Length 2874;  
Best Local Similarity 76.0%; Pred. No. 3.5e-298;  
Matches 1778; Conservative 0; Mismatches 266; Indels 297; Gaps 3;  
QY 1303 GCGGCAAAACGGGAACTTGAAGGCAACGCACTTGAATGGGAACGGACCGGAGACAG 1362  
DB 1 GCTCAAAACGGGAACTTGAAGGCAACGCACTTGAATGGGAACGGATCGAAGGCAA 60  
QY 1363 GAACCTCTGAATCAGAGAAACAAGAGCAGAGGAGCACCCTGGTCTGAAGCAGAGAGC 1422  
DB 61 GAACCTCTGAATCAGAGAAACAAGAGCAGAGGACATAGTTGATCTGAAGCAAGAA 120  
QY 1423 AAGACTCTGAGTTTGAATGAGAAAGCTCTGAATGACAAAAAGCATCAGTAAGCAAAA 1482  
DB 121 AAGACTTTGGAATTTGAATTAAGAAAGCTTAATGATTAAGAAAGCATCAATGAAGGAAA 180  
QY 1483 CTTGAGATATCAGGTGCTGAGCTGAGCAACCCAGAGGCAAAATTTGAGAGCAAGAACAG 1542  
DB 181 CTTGAGATATCAGATGTGATGACCAACCAAGGCAAAATTTGAGAGCAACAAACAA 240  
QY 1543 TCTAGAGAGCTTAAGAAATTTGCTGAATCACCACCTTACAGCAGAGTTCAGAAATCTCAG 1602  
DB 241 TCTAGAGAGTTGAGAAATTTGCTGAATCACCACCTTACAGCAGAAATTTACAGAAATCTCAG 300  
QY 1603 CAAATGCTTGAAGACTTATTTCCAGAGAAACAGATACCTAGTACCAAGTTAAACAAATC 1662  
DB 301 CAAATGCTTGAAGACTTATTTCCAGAGAAACAGATACCTAGTACCAAGTTAAACAAATC 360  
QY 1663 CAGAGAAAGCTTTGCACTAAGAGCTGCTTCACTTCACTTCACTTCACTTCACTTCACTT 1722  
DB 361 CAGAGAAAGCTTTGCACTAAGAGCTGCTTCACTTCACTTCACTTCACTTCACTTCACTT 420  
QY 1723 GAGCTGCCCCGAGCAGCTCCGAGAGCAGCTGAGCAGAGTGGAGAGAGAGCAGAGTCA 1782  
DB 421 GAACTAGCTGGGAGCAGCTTACAGAGACCACTGATGATGAGAGAGAGAGAGAGTCA 480  
QY 1783 AAGCTGAGAGATTTGATTTTCAACCAACCACTGAGAGAGAGAGATGATGATGATGATG 1842  
DB 481 AACTTACAGAGATTTGATTTTCAATATATCAGTGAAGAGATTAAGAGAAATATACCAAT 540  
QY 1843 AAACAGCACTCCAGAGAGAGAGTCCCTGAGAGCAGCGGACTGAAGCAGAAAGAGCAG 1902  
DB 541 AAGCAACAACTCCAGAGAGAGAGTCCATGAGAGCTGAAGAGCTGAAGAGAGAGAGCAG 600  
QY 1903 GAGAGAGAGAGCTGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1962  
DB 601 GAGAGAGAGAGATTTGATTTTCAACCAACCACTGAGAGAGAGAGAGAGAGAGAGAGAG 660  
QY 1963 AAGGACAAAGAGATTTGATTTTCAACCAACCACTGAGAGAGAGAGAGAGAGAGAGAG 2019  
DB 661 AAGGACAAAGAGATTTGATTTTCAACCAACCACTGAGAGAGAGAGAGAGAGAGAGAGAG 720  
QY 2020 CCCCACGAG 2079  
DB 721 CTCACAG 780  
QY 2080 AGAGCAG 2139  
DB 781 AAGGACAAAG 840  
QY 2140 GCTAAGCTGGCCACCCAGAGACCTGCTTACCAAGAGAGAGAGAGAGAGAGAGAGAGAG 2199





CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

CC Sequence 2131 BP; 691 A; 448 C; 572 G; 420 T; 0 other;

Query Match 34.6%; Score 1259.2; DB 22; Length 2131;

Best Local Similarity 81.2%; Pred. No. 1.5e-290;

Matches 1539; Conservative 0; Mismatches 243; Indels 114; Gaps 2;

QY 1 AAGGCTCAGTTTCCACACCTTTCGGTAGCTGAGCTGGGCCCTTAACCTGGAG 60  
 DB 347 ATGGCTCAGTTTCCACACCTTTCGGTAGCTGAGCTGGGCCCTTAACCTGGAG 406  
 QY 61 GAAAGGCCAAGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120  
 DB 407 GAAAGGCCAAGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 466  
 QY 121 GGTGATCAAGCCGAGAACTTTTTCATCTGGGTACCTGAGCTTCTTAGACAA 180  
 DB 467 GGTGATCAAGCCGAGAACTTTTTCATCTGGGTACCTGAGCTTCTTAGACAA 526  
 QY 181 ATATGGGCGCTAGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 240  
 DB 527 ATATGGGCGCTAGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 586  
 QY 241 GCCATGAAGCTTATCAAACTGAAAGTACAAAGATATCACTCCCTCCACACTTCCCT 300  
 DB 587 GCCATGAAGCTTATCAAACTGAAAGTACAAAGATATCACTCCCTCCACACTTCCCT 646  
 QY 301 GTCATGAACAGCAACCACTGCTATTTCCAGTCAACCACTTGGTATGAGAGGATT 360  
 DB 647 GTCATGAACAGCAACCACTGCTATTTCCAGTCAACCACTTGGTATGAGAGGATT 691  
 QY 361 GGTAGATGCAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420  
 DB 692 ----- 691  
 QY 421 GGAATGTCACACCTTAGATCTCTGCTCCCTCAGACAGAGTGCCTCCCTGGCTAAC 480  
 DB 692 ----- GCAAGCTGTGCCCCCTGGTAAAC 715  
 QY 481 GGGGCTTCCTCCGTCATACAGCTTGGCTTGGCGATCTGACACCAATGGCCA 540  
 DB 716 GGGGCTTCCTCCGTCATACAGCTTGGCTTGGCGATCTGACACCAATGGCCA 775  
 QY 541 AAGAGTTCCTCCGTCATACAGCTTGGCTTGGCGATCTGACACCAATGGCCA 600  
 DB 776 AAGAGTTCCTCCGTCATACAGCTTGGCTTGGCGATCTGACACCAATGGCCA 835  
 QY 601 GCAACAATCAATTCGATGCGCAGCGCCCTCCAGCAGCAATGGCTGTGCTGAGTCA 660  
 DB 836 GCAACAATCAATTCGATGCGCAGCGCCCTCCAGCAGCAATGGCTGTGCTGAGTCA 895  
 QY 661 TCAAGGCTGAATATACAGGCACTTATTCACAGCCACGACAAAATCTATGAGTGACACTTA 720  
 DB -----

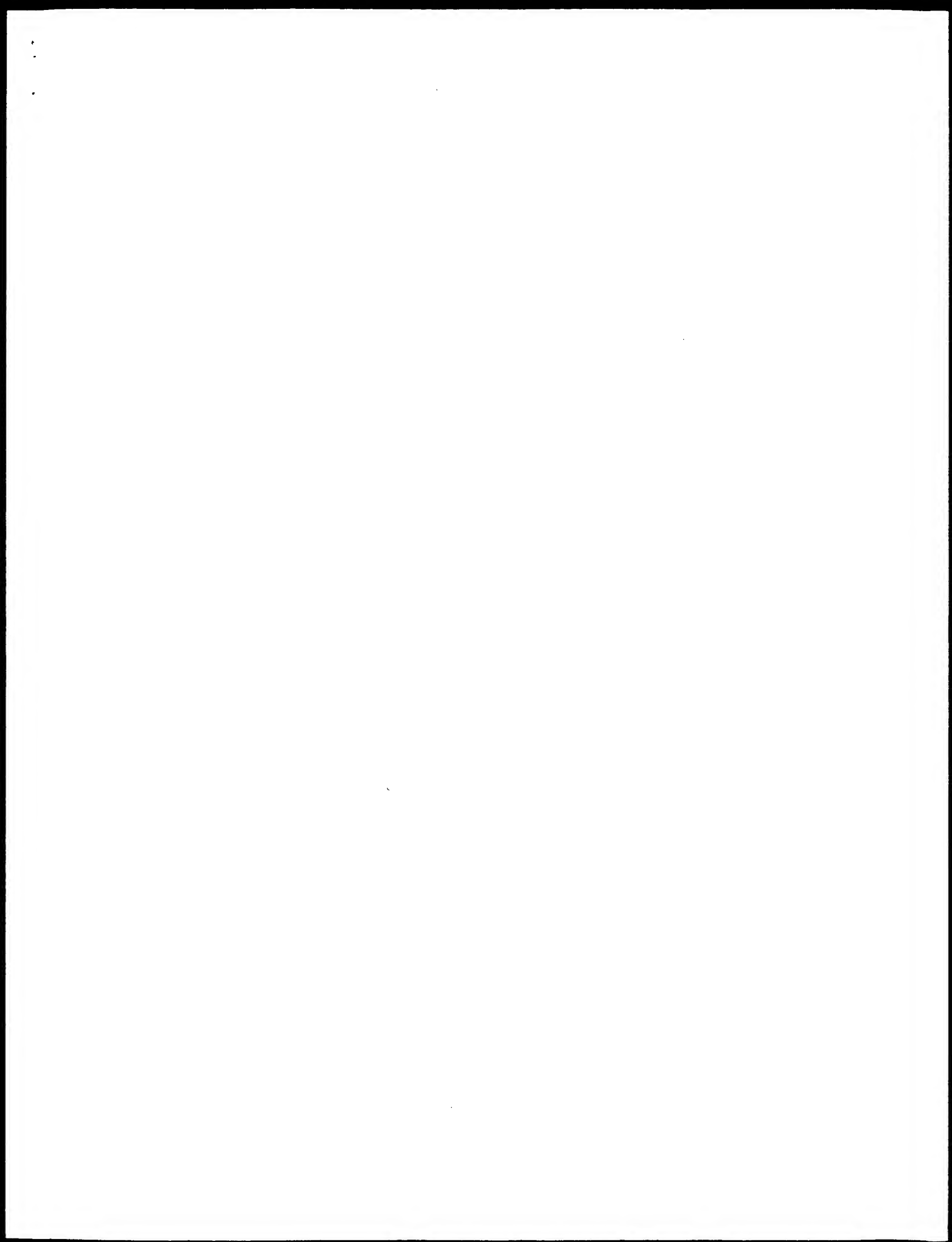
DB 896 TCAAGCTGAATATACAGGCAATTAATTCATATGATGACAAAATCTATGAGTGACACTTA 955  
 QY 721 ACAGGTCCTCCAGCAAGAACTATTTCTATGATCAATCAAGTTTACCCAGGCTCAGTGCT 780  
 DB 956 ACAGGTCCTCCAGCAAGAACTATTTCTATGATGATCAAGTTTACCAAGGCTCAGTGCT 1015  
 QY 781 TCAATATGAAATCTTTCTGACATTTGATCAAGATGAAAACTCACTGACAGAAATTTATC 840  
 DB 1016 TCAATATGAAATCTTTCTGACATTTGATCAAGATGAAAACTCACTGACAGAAATTTATC 1075  
 QY 841 CTACATATGACCACTTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900  
 DB 1076 CTGCAATGACCTCATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1135  
 QY 901 CCAAAATACATCCCT 960  
 DB 1136 CCAAAATACATCCCT 1195  
 QY 961 AGCTCTTCT 1020  
 DB 1196 AGCTCTTCT 1255  
 QY 1021 C---CAGAGAAAGAACTCCCTGTCATCTTGAAGATGAAAGGCGGAGAACTTCAGCGA 1077  
 DB 1256 CAATTGAAAGAAATTAATCTGTAACCTTGAAGATGAAAGGCGGAGAACTTTGAACGT 1315  
 QY 1078 GGCAGTGTGAGCTGAGAGAGCGCCGCAAGCGCTCTTGGAGCAGCAGCCCAAGAGCAG 1137  
 DB 1316 GGCAGTGTGAGCTGAGAGAGCGCCGCAAGCGCTCTTGGAGCAGCAGCCCAAGAGCAG 1375  
 QY 1138 GAGCGTTTGTCTGCTGAGAGCGCGCCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1197  
 DB 1376 GAGCGTTTGTCTGCTGAGAGCGCGCCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1435  
 QY 1198 CAGAGGCGCAAGCGCGAGCTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1257  
 DB 1436 CAGAGGCGCAAGCGCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1495  
 QY 1258 CGGAGCGAG 1317  
 DB 1496 CAGAGCGAG 1555  
 QY 1318 CTGAAAGCGAGCGCAACTTGAATGGGAAAGGAAACCGGAGACAGAACTCTGAATCAG 1377  
 DB 1556 CTGAAAGCGAGCGCAACTTGAATGGGAAAGGAAACCGGAGACAGAACTCTGAATCAG 1615  
 QY 1378 AGGAAACAG 1437  
 DB 1616 AGGAAACAG 1675  
 QY 1438 GAGTTAGAGCTGTGATGACAAAGAAACATCAGTGAAGAGAAACCTTCAGATATCAGG 1497  
 DB 1676 GAGTTAGAGCTGTGATGACAAAGAAACATCAGTGAAGAGAAACCTTCAGATATCAGG 1735  
 QY 1498 TGTGACTGGCAACCCAGAGGCAAGAAATTTAGAGACGAAAGTCTAGAGAGCTAAGA 1557  
 DB 1736 TGTGACTGGCAACCCAGAGGCAAGAAATTTAGAGACGAAAGTCTAGAGAGCTAAGA 1795  
 QY 1558 ATTGCTGAATACACCACTTACAGCAGCAGTGTGAGAGAACTCAGCAAAATGCTTGAAGA 1617  
 DB 1796 ATTGCTGAATACACCACTTACAGCAGCAGTGTGAGAGAACTCAGCAAAATGCTTGAAGA 1855  
 QY 1618 CTATATTCAGAGAAAGATTAATCAGTGAAGCAGTGAAGCAAGTCCAGCAGAAAGCTTTG 1677  
 DB 1856 CTATATTCAGAGAAAGATTAATCAGTGAAGCAGTGAAGCAAGTCCAGCAGAAAGCTTTG 1915  
 QY 1678 CATAGAGCTGCTTCTTACCTCAAAAGAGCTTGGAGCAAGAGAGCTGCGCCGAG 1737  
 DB 1916 CATAGAGCTGCTTCTTACCTCAAAAGAGCTTGGAGCAAGAGAGCTGCGCCGAG 1975  
 QY 1738 CAGTCCGGAGAGAGCTGAG 1797  
 DB 1976 CAGTCCGGAGAGAGCTGAG 2035



Qy 2403 GGTTCCTCCTCCAGCCAAACCACTGACATCTGACATCTGCCCCCTGCCCCCAAACTGGC 2462  
 Db 901 GGTTCCTCCTCCAGCCAAACCACTGACATCTGACATCTGCCCCCTGCCCCCAAACTGGC 960  
 Qy 2463 TCTGCGTGAAGACCCCTGCTCTTTGCGAGTGAACCTTTGAGCCCTCCAGACCCCA 2522  
 Db 961 CTTCGCTGAAGACCCCTGCTCTTTGCGAGTGAACCTTTGAGCCCTCCAGACCCCTNA 1020  
 Qy 2523 CAACTGGGAGACCTTCAGTTCCACGTGGCCCAAGCAGCTCAAAAGAAAGCCAGAAACGGA 2582  
 Db 1021 TAACTGGGCGCACTTCAGTTCCACGTGGCCCAAGCAGCTCAAAAGAAAGCCAGAAACGGA 1080  
 Qy 2583 CAACTGGGATACGTGGGCGGCTCAGCTTCTCTGACCCGTAAGTGTGCTGGCCAGTTAGC 2642  
 Db 1081 TAACTGGGATACGTGGGCGGCTCAGCTTCTCTGACCCGTAAGTGTGCTGGCCAGTTAG 1140  
 Qy 2643 GAGAGATCAGCTTTACCCAGCCAGCCAGCCAGCTGCTCCCAATCCCGCTGGG 2702  
 Db 1141 GAGAGATCAGCTTTACCCAGCCAGCCAGCCAGCTGCTCCCAATCCCGCTGGG 1200  
 Qy 2703 CCAGGCTGAAAAGTGAAGGGCTACAGCCAGCCAGCTGTATCCCTGGAGAGCCAAAA 2762  
 Db 1201 CCAGGCTGAAAAGTGAAGGGCTACAGCCAGCCAGCTGTATCCCTGGAGAGCCAAAA 1260  
 Qy 2763 AGACAACTCACTTAATTTTAAACAAAAGTACGTCACTCAGCTTCTGGAACAGCAAGACAT 2822  
 Db 1261 AGACAACTCACTTAATTTTAAACAAAAGTACGTCACTCAGCTTCTGGAACAGCAAGACAT 1320  
 Qy 2823 GTCGCTGTTGAGAGTTCAAGTGCAGAGGGTTCCTCCCAAGCTTACGTGAACAT 2882  
 Db 1321 GTCGCTGTTGAGAGTTCAAGTGCAGAGGGTTCCTCCCAAGCTTACGTGAACAT 1380  
 Qy 2883 CATTCAGGGCCGTTAAGAAATCCACAGCATCGATCTGCGCCCTACTGAAAAGTCTGC 2942  
 Db 1381 CATTCAGGGCCGTTAAGAAATCCACAGCATCGATCTGCGCCCTACTGAAAAGTCTGC 1440  
 Qy 2943 TAGCTTAAAGAGAGTGGCTTCCCGGCGCCCAAGCCAGTTCGCGAGAAAGTTAT 3002  
 Db 1441 TAGCTTAAAGAGAGTGGCTTCCCGGCGCCCAAGCCAGTTCGCGAGAAAGTTAT 1492  
 Qy 3003 TGCATGTACACATACAGAGTTCTGAGCAAGAGATTAACTTTACGAAAGGGATGT 3062  
 Db 1493 ----- 1492  
 Qy 3063 GATTGTGTTACCAAGAAAGATGTGACTGTGACGGGACGGTGGCGACAAGTCCGG 3122  
 Db 1493 ----- 1492  
 Qy 3123 AGTCTTCCTCTTAATCTATGTAGGCTTAAAGATTGAGAGGGCTTGAACTGTGGGA 3182  
 Db 1493 ----- 1492  
 Qy 3183 AACAGGAGATTAGAAAAAAACCTGAAATTGCGCAGTTATTGCTTCTACGCTGTAC 3242  
 Db 1493 -----GAAATTGCGCAGTTATTGCTTCTACGCTGTACGCGCAC 1527  
 Qy 3243 TGGTCCCGAACAACCTGAGCTCTGCGGAGCTGATTCGATCCGGAAAAAGAACCC 3302  
 Db 1528 CGGCCCCGAGACACTCACTTCGCCCCCTGCTGAGTTTGTATCGAAAAAGAACCC 1587  
 Qy 3303 AGGTGATGTGGAGAGAACTGCAAGCTCGAGGAAAAAGCGCAGATAGGTGTT 3362  
 Db 1588 AGGTGATGTGGAGAGAACTGCAAGCTCGAGGAAAAAGCGCAGATAGGTGTT 1647  
 Qy 3363 TCCAGCAATTTATGTCAAACTTTAAAGCCCCGGAACAGCAAAATCACTGAGCT 3422  
 Db 1648 CCCAGCTAATTTATGTAAAGCTTCTAAAGCCCTGGGACGAGCAAAATCACTCAACAGAGCC 1707  
 Qy 3423 ACCAAGACCGGAGTGCAGCAGTGCAGAGTGCAGAGTGCAGAGTGCAGATTAACACCGC 3482  
 Db 1708 ACCTAAGTCAACAGCATTAAGGCGAGTGCAGAGTGCAGAGTGCAGATTAACACCGC 1767  
 Qy 3483 CCAGAAAGATGAGAACTAGCTTTCAGCAAAAGCCAGATCATCACTCTCAACAGGA 3542

Db 1768 GCAGAAATGACATGAGCTGCGCTTCAACAAAGGCGCAGATCATCACTCTCAACAAAGGA 1827  
 Qy 3543 GACCCGAGCTGTGGAAGAGAGTCAAGTGGGCAAGTTGGGCTCTTCCATCAATTA 3602  
 Db 1828 GACCCGAGCTGTGGAAGAGAGTCAAGTGGGCAAGTTGGGCTCTTCCATCAATTA 1887  
 Qy 3603 TGTAAAGCTGACACAGACATGGAACCCAGCCAGCAATGA 3642  
 Db 1888 TGTAAAGCTGACACAGACATGGAACCCAGCCAGCAATGA 1927

Search completed: March 14, 2003, 02:10:36  
 Job time : 699.685 secs







Db 402 ATGCCAATCTGTCATTCATCAAGCCATGCTCCATTCAGCCATTAAGCAACACCTTG 461  
Qy 401 -----TGGCTCCATTCAGTTGTTGAATGTCTCCACCTTAGTATCT 444  
Db 462 TCTTCTGCTACTTCAGGAGCAGATATTCCTCCCTAATGATGCTGCTCCCTAGTCCCT 521  
Qy 445 TCTGCTCCCTCAGCAGAGTGGCTCCCTGCTTAACGGGGCTCTCCCTCATACAGCCT 504  
Db 522 TCTGTAGTACATCTCTATTAACAAATGGAATGCCAGTCTCATTCAGCCTTATCCATT 581  
Qy 505 CTGCTGCTGTTGGCAGTCTGACCAATGCGCAAAAGTTCTCTTCAAGATCT 564  
Db 582 CTTATTTCTTCTCAACATGCTCATGACATCTTACAGCCTGATGATGAGGAGATT 641  
Qy 565 GGTCCAGGTCACAAATTAACATAGTTACAGAGGCAATATTCATGTCGCGAGC 624  
Db 642 GGTGTGCTAGTATCCAGAAAGCCAGTCTCTGATGATTTAGGATCTAGTACTCACT 701  
Qy 625 GCCCTCCAG-----CAGCAGATGGGCTGTG 651  
Db 702 TCCCTCACTGCTTCCCTCTCAGGGACTCACTAAGACAGGACCTCAGAGTGGGAGTT 761  
Qy 652 CCTAGTATCAGGCTGAATACAGGCAATTAATCAACGCCACGCAAACTATGACT 711  
Db 762 CTTAGGCTTCAGATTAAGATAGGCAAAATTTAATAGTCTAGCAAAAGGCAATGAGC 821  
Qy 712 GACACTTAACAGTCCCGAGGCAAGAACTATTCTCATGCAATCAATTAACCCAGGCT 771  
Db 822 GGAATCCTCTCAGGTTTTCAGAGCTGAAATGCCCTTCTTCACTCAATCTCTCTCAACT 881  
Qy 772 CAGTGCCTTCATATGAGATCTTCTTGAATGATGATCAAGATGAGAACTCAGTCAGAA 831  
Db 882 CAGTACTACTATTTGACTCTGCTGACATGATGATGATGATGATGATGATGATGATGAT 941  
Qy 832 GAATTTATCTAGCTATGACCTTAATGATGATGATGATGATGATGATGATGATGATGATGAT 891  
Db 942 GAATTTATCTGAGATGACCTCAGTCACTGACATGATGATGATGATGATGATGATGATGAT 1001  
Qy 892 GTCTGCTCTCAGAAATCATCT 951  
Db 1002 ACGTTGCTCCGAGCTTGTCT 1050  
Qy 952 TCCCTCATAGCT 1011  
Db 1051 -----TTGATCTGTTAATGAACTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1097  
Qy 1012 GAGCAGCAGCCAGAGAAAGAACT 1071  
Db 1098 GA---AGAGCTCAGAAAGAACT 1154  
Qy 1072 GAGGAGGCACTGTGAGCTGTGAGAGAGCCGCGCAAGCGCTCTTGGAGCGAGCGCAAA 1131  
Db 1155 GAAAGAGAAACATGAGCTGTGAGAGAGAGCCCAAGTGTGATGAGAGCGAGCAGAGG 1214  
Qy 1132 GAGCAGGAGCGGTTGCTCAGCTGAGCGCGCCGAGCAGAGAGAGAGAGAGAGAGAGAGAGAG 1191  
Db 1215 GAGGCTGAACGCAAG 1274  
Qy 1192 CAGAGCAGAGAGCGCAAGCGGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1251  
Db 1275 CAGAGCAG 1334  
Qy 1252 CTGAGAGGAG 1311  
Db 1335 CTGAG 1394  
Qy 1312 CGGAACTGGAAG 1371  
Db 1395 CAGAGAGCTTGAAG 1454  
Qy 1372 AATCAG 1431

Db 1455 AGTCAGAGACCAGGAGCAAGAGACATTTGTCAAGCTGAGCTCCAGAAAGAAAGCTCTC 1514  
Qy 1432 GAGTTGAGTTAGAAAGCTCTGAATGAACAAAGAGATCAGCTAGAGAGAGAGAGAGAGAGAG 1491  
Db 1515 CAGCTGGAACGTGAAG 1574  
Qy 1492 ATCAGAGTGCAGTGGCAACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1551  
Db 1575 GTCCAAATCAGAAAGCAAAAG 1634  
Qy 1552 CTAAAGATTGCTGAATATCCACTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1611  
Db 1635 CTGAAATTAATGAAATTAACCACTTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1694  
Qy 1612 GGAAGACTTAATCCAG 1671  
Db 1695 ATCTATCTGCTCCTGAG 1754  
Qy 1672 AGTTGATAG 1731  
Db 1755 AACACAGCTGATTCAG 1814  
Qy 1732 CGGAG 1791  
Db 1815 TGCCAAAG 1874  
Qy 1792 GAGATTAATGTTTTCACACAG 1851  
Db 1875 GAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1934  
Qy 1852 CTCAG 1911  
Db 1935 TTAGGCTTTGAACATTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1994  
Qy 1912 AGCTGAGAGTAG 1934  
Db 1995 AGATTAG 2017

RESULT 2  
US-08-630-915A-193  
; Sequence 193, Application US/08630915A  
; Patent No. 6309820  
; GENERAL INFORMATION:  
; APPLICANT: SPARKS, Andrew B.  
; APPLICANT: HOFFMAN, No. 6309820h  
; APPLICANT: KAY, Brian K.  
; APPLICANT: FOWLES, Dana M.  
; APPLICANT: McCONNELL, Stephen J.  
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL  
; TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
; NUMBER OF SEQUENCES: 227  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/630,915A  
; FILING DATE: 03-APR-1996  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mirock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 1101-174

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 790-9090  
 TELEFAX: (212) 869-8864/9741  
 TELEX: 66141 PENNIE  
 INFORMATION FOR SEQ ID NO: 193:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2873 bases  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: unknown  
 MOLECULE TYPE: DNA  
 US-08-630-915A-193

Query Match 9.1%; Score 332.4; DB 4; Length 2873;  
 Best Local Similarity 61.0%; Pred. No. 1.9e-73;  
 Matches 631; Conservative 0; Mismatches 361; Indels 42; Gaps 4;

QY 2609 CTCTCTGACCGTACTAGTCTGGCCAGTTCAGGACGAGATAGCCTTTACCCCGCA 2668  
 DB 541 CTTTTTCAAACTTAACCTTAATAATCATCATGCGCAAAAAATCAGCTTCACTCGAACTG 600  
 QY 2669 CAGCCACTGGCTCTCCCATCTCCCGTCTCGGCGGAGTGAAGTGAAGGCTAC 2728  
 DB 601 TGTCCCTG--ATCTGATCACTATTCATGACACAGGACAAAGTGAAGAACTTAA 657  
 QY 2729 AAGCGCAAGCCTGTATCCCTGAGAGCCAAAAAGACCACTTAAATTTTACAAAA 2788  
 DB 658 AAGCAGAGGCTTTGTCTCTGACAGCAAGAAAGATTAACCTTGAACCTCTCAAAAC 717  
 QY 2789 GTACGCTACACCTTTCTGGAACAGCAAGACATGTGTGTTTGAAGAAGTTCAAGTTC 2848  
 DB 718 ATACATATTAATGTTGTTGAGACAGCAAGAAATTTGGTGGAGGTGATGAGAG 777  
 QY 2849 AGAAGGTTGTTCCCAAGTCTTAAGTCAATTTCAAGGCGCCGTAAGAAATCA 2908  
 DB 778 GAAGAGATGTTTCCCAATCTTAATGTCAAGATCATTTCTGGAGTGAAGTAAACGGG 837  
 QY 2909 CAGCATTCGACTGCGCTTAAAGTCTCTAGTCTAAAGAGATGGCTTCCCGG 2968  
 DB 838 AA-----GAACCAAGAGCTTTGATGACAGCTGTAATTAAGAAACCTACCTCGG 885  
 QY 2969 CCGCCAAAGCAGCCATTCGCGGAGAGATTTATGCGATGTACATACAGAGATTTCTG 3028  
 DB 886 CAGCTTATTCAGTT-----GGAGAAAGATATATGACCTTTATCCATATTCAGAGTGTG 939  
 QY 3029 AGCAGAGATTAACCTTACAGCAAGGAGATGATGTTGTGTTACCAAGAAAGATGTG 3088  
 DB 940 AACCTGAGATTTGACTTTCACAGAAAGTGAAGAAATATTTGTGACCCGAAAGAGTGGAG 999  
 QY 3089 ACTGTGAGCGGAAACGATGCGGACAAAGTCCGAGTCTTCCCTTAACTATGTAGAGC 3148  
 DB 1000 AGTGTGAGAGAGATTTGAGATGAGATGAGTGAATTTTTCATCAACTATGTCAAAAC 1059  
 QY 3149 TTAAGATTCAGAGGCTGTGAACTGCTGGGAAAAACAGGAGATTAGAAAAAAACCTG 3208  
 DB 1060 CAAAGGATCAAGAGAGATTTGGAGTGTCTAGCAAGTCTGAGCATCAATTAATAAACCTG 1119  
 QY 3209 AAATTCGCGAGTTATGCTTCTTAAGCTCTAAGTCTGTTCCCGAAACATCACCTGGCTC 3268  
 DB 1120 AGATGCTCAGTAACTTCAGCATATGTTCTTGTGTTTGAACAACCTTAAGCTTGGAC 1179  
 QY 3269 CTGGGACGATTTCTGATCCGAAAAAGAACCCAGGTGATGTTGGGAAAGAGAACTGC 3328  
 DB 1180 CAGGACAGTTAATTAATTAATTTAAAGAAAAATACAAAGTGGTGTGTCAGAAAGAGATTAC 1239  
 QY 3329 AAGCTGAGGAAAAAGCCGCAAGTAGGTGTTTCCAGCAAAATATATGCAAACTTTCAA 3388  
 DB 1240 AGGCAAGAGAAAAAGGAGCAAGAAAGATGTTTCTGCGCACTCATGTTAACTTTTGG 1299  
 QY 3389 GCGCCGGAACAGCAAAATACCCCACTAGCTACCCCAAGACCGAGTGCACGCGAG 3448  
 DB 1300 GTCCAAAGTATGAAGGACCACTGCTTTTATC-----CTG 1338

QY 3449 TGTGCGAGTATGGGATGTAGCATTAACCCCGCAAGCGATGAGAACTAGCCTTCA 3508  
 DB 1339 TATGTCAGTATGCTATGATATGATGACAAATTAATGAAGATGACTCAGTTCT 1398  
 QY 3509 GCAAAGCCAGATCATCAAGCTCCTCAACAAGAGACCCGAGCTGTGGAAAAAGAGAG 3568  
 DB 1399 CCAAGGACAACTATTAATTTATGAACAAAGATGATCTGATGTGTGCAAGAGAGA 1458  
 QY 3569 TCAGTGGCAAGTTGGGCTTTCCCATCAATTATGTAAAGCTGACCAAGACATGAGC 3628  
 DB 1459 TCAAGGGGTGACTGTGCTCTTCTTCAAACTAGCTTAAGATGACAGACATCAGATC 1518  
 QY 3629 CAGCCAGCAATGA 3642  
 DB 1519 CAGTCAACAGTGA 1532

RESULT 3  
 US-08-630-915A-39  
 Sequence 39, Application US/08630915A

Patent No. 6309820  
 GENERAL INFORMATION:  
 APPLICANT: SPARKS, Andrew B.  
 APPLICANT: HOPMAN, No. 6309820h  
 APPLICANT: KAY, Brian K.  
 APPLICANT: FOWLES, Dana M.  
 APPLICANT: McCONNELL, Stephen J.  
 TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL  
 TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
 NUMBER OF SEQUENCES: 227  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: Pennie & Edmonds LLP  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10036-2711

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/630,915A  
 FILING DATE: 03-APR-1996  
 CLASSIFICATION: 536  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Mistrock, S. Leslie  
 REGISTRATION NUMBER: 18,872  
 REFERENCE/DOCKET NUMBER: 1101-174  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 790-9090  
 TELEFAX: (212) 869-8864/9741  
 TELEX: 66141 PENNIE  
 INFORMATION FOR SEQ ID NO: 39:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 747 bases  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA  
 US-08-630-915A-39

Query Match 7.1%; Score 258; DB 4; Length 747;  
 Best Local Similarity 63.9%; Pred. No. 4.1e-55;  
 Matches 418; Conservative 0; Mismatches 215; Indels 21; Gaps 1;

QY 2989 GGAGAGAGTTATGCGCATGACATACAGAGATTCTGAGCAAGAGATTAACTTAC 3048  
 DB 115 GGAGAGATATATTTGCACTTATTCATATTCAGTGTGAGAACTGAGATTTGACTTTC 174  
 QY 3049 CAGCAAGGAGATGTGATTTGTGTTACCAAGAAAGATGTGACTGTGTGACGGAAAGG 3108

```

Db      175  ACAGAAAGTGAAGAAATATTGTGTACCCAGAAAGATGAGAGTGAGACAGAAAGATT 234
QY      3109  GGGACACAGTCCGAGTCTTCCCTTCTACTAGTGTAGGGTTAAAGTTCAAGGGGCTC 3168
Db      235  GGAGATGAAAGTGAATTTTTCATCAAACTATGTCAAACCAAGATCAAGAGATTTT 294
QY      3169  GGAATGCTGGGAAACAGAGGAGTTTAGAAAAAACCTGAATTGGCCAGGTTATTGCT 3228
Db      295  GGGAGTGTAGCAAGTCTGGAGCATCAATATAAAAAACCTGAGATGTGCTCAGTAACCTCA 354
QY      3229  TCCAGCTGTACTGTGTCCCGAACAACCTCAGCTGTGCTGGGAGCTGATTCGATC 3288
Db      355  GCATATGTGTCTGTGTCTGAACAACCTTAGCTTGCACAGAGACGTTAAATTAATT 414
QY      3289  CGGAAAAAAGAACCCAGGTGTAGTGTGGAAAGAGAACTGCAAGCTCAGAGGAAAAAGCCG 3348
Db      415  CTAAGAAAAAATACAAATGGGTGTGGCAAGAGAGTTACAGGCAAGGAAAAAGCCGA 474
QY      3349  CAGATAGGGTGTGTCCAGCAAAATTATGTCAAACTTCTAAGCCCGAGAACCAAAATC 3408
Db      475  CAGAAAGATGTGTCTTCCGACATCATGTTAACTTTGGGTCCAGACAGTGAAGAGCC 534
QY      3409  ACCCCACTGAGCTAACCCAGACCGCAGTGTGAGCCAGAGTGTGCGGATG 3468
Db      535  ACACCTGCTTTTCTCTCT-----GTATGTCAAGTGAATTGCTATG 573
QY      3469  TACGATTACACCGCCAGAAAGATGACGAACCTAGCTTACGAAAGGCGCATCATCAAC 3528
Db      574  TATGACTATGAGCAAAATATGAAGATGAGCTCAGTTTCTCAAGGAGCAACTCATTAAT 633
QY      3529  GTCTCAACAGAGGAGACCGGAGCTGTGTGAAAGAGAAAGTCAAGTGGGCAAGTTGGGCTC 3588
Db      634  GTATGAAACAAAGATGATCTGATTTGTGTGCAAGAGAGATCAACGGGGTGACGTGCTC 693
QY      3589  TTCCATCATCATTTATGTAAGCTGACCAAGACATGAGACCCGACGCAATGA 3642
Db      694  TTTCCTCAAACTAGTTAAGATGACGACAGCTCAAGTCAACAGTGA 747

```

```

RESULT 4
US-09-404-879A-60
; Sequence 60, Application US/09404879A
; Patient No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 60
; LENGTH: 480
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-404-879A-60

```

```

Query Match      4.0%; Score 144; DB 4; Length 480;
Best Local Similarity 62.1%; Pred. No. 1.1e-26;
Matches 247; Conservative 0; Mismatches 145; Indels 6; Gaps 1;

```

```

QY      1  ATGGCTCAGTTTCCCAACCTTTGGTGTGACCTGATGTCTGGCCATTAACCTGTGGAG 60
Db      14  ATGGCTCAGTTTCCCAAGCGATGAATGAGGGCCAAATATGTGGCTATTACATCTGAA 73
QY      61  GAAAGGGCAAGCATGACGACGATTCCTTAGCCTGAAGCGAATAGCGGAGTTATTACT 120
Db      74  GAAAGTACTAAGCATGATTAACAGATTGTATTAACCTCAAACTTCAGAGGTTACATACA 133

```

```

QY      121  GGTGATCAAGCGAGAACTTTTTCCTCAATCTGGTTACTCAGCTGTCTTAGCAAA 180
Db      124  GGTGATCAAGCCCGTACTTTTTCCTCAGCAGAGTGTGCGGCGCGGTTTAGCTGAA 193
QY      181  ATATGGGCGCTAGGGAATGAATTAAGATGGAAGATGATCAAGTGAATTTCCATA 240
Db      194  ATATGGGCTTATCAGATCTGAACAGAGATGGAATGGAACAGCAAGATTTCTATA 253
QY      241  GGCATGAAGCTTATCAAACTGAAGCTACAAAGATATAGCTCCCTCCACACTTCCCT 300
Db      254  GGTATGAAGCTCATCAAGTTAAAGTTGAGGGCCAAACAGCTGCTGTATGTCCTCTCT 313
QY      301  GTATGAAGAACAGCAACAGTGTCTATTTCCAGTGCACCAGATTTGGTATAGAGGAT 360
Db      314  ATCATGAAGAACACCC-----CTATGTCTCTCCTCACTAATCTCTGCTTTGGGATG 367
QY      361  GCTAGCATGCACACCACTCAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 398
Db      368  GGAAGCATGCCCATCTGTCTCATTCATTCAGCATTTGCC 405

```

```

RESULT 5
US-09-404-879A-5
; Sequence 5, Application US/09404879A
; Patient No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 531
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-404-879A-5

```

```

Query Match      4.0%; Score 144; DB 4; Length 531;
Best Local Similarity 62.1%; Pred. No. 1.2e-26;
Matches 247; Conservative 0; Mismatches 145; Indels 6; Gaps 1;

```

```

QY      1  ATGGCTCAGTTTCCCAACCTTTGGTGTGACCTGATGTCTGGCCATTAACCTGTGGAG 60
Db      42  ATGGCTCAGTTTCCCAAGCGATGAATGAGGGCCAAATATGTGGCTATTACATCTGAA 101
QY      61  GAAAGGGCAAGCATGACGACGATTCCTTAGCCTGAAGCGATAGCGGAGTTATTACT 120
Db      102  GACGTAATCAAGCATGATTAACAGTTTGATTAACCTCAAACTTGAAGAGTTACATAACA 161
QY      121  GGTGATCAAGCGAGAACTTTTTCCTCAATCTGAGTTACTGAGCTGTCTTAGCAAA 180
Db      162  GGTGATCAAGCCCGTACTTTTTCCTCAGCAGAGTGTGCGGCGCCGGTTTAGCTGAA 221
QY      181  ATATGGGCGCTAGGGAATGAATTAAGATGGAAGATGATCAAGTGAATTTCCATA 240
Db      222  ATATGGGCTTATCAGATCTGAACAGAGATGGAAGATGGAACAGCAAGATTTCTATA 281
QY      241  GGCATGAAGCTTATCAAACTGAAGCTACAAAGATATAGCTCCCTCCACACTTCCCT 300
Db      282  GGTATGAAGCTCATCAAGTTAAAGTTGAGGGCCAAACAGCTGCTGTATGTCCTCCT 341
QY      301  GTATGAAGAACAGCAACAGTGTCTATTTCCAGTGCACCAGATTTGGTATAGAGGAT 360
Db      342  ATCATGAAGAACACCC-----CTATGTCTCTCCTCACTAATCTCTGCTTTGGGATG 395
QY      361  GCTAGCATGCACACCACTCAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 398
Db      368  GGAAGCATGCCCATCTGTCTCATTCATTCAGCATTTGCC 433

```

## RESULT 6

US-08-728-323A-1  
 ; Sequence 1, Application US/08728323A  
 ; Patent No. 5948676

## GENERAL INFORMATION:

APPLICANT: Chang, Yuan  
 APPLICANT: Bohenzky, Roy A.  
 APPLICANT: Russo, James J.  
 APPLICANT: Edelman, Isidore S.  
 APPLICANT: Moore, Patrick S.  
 TITLE OF INVENTION: Immediate Early Protein From Kaposi's  
 TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA  
 NUMBER OF SEQUENCES: 21  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Cooper & Dunham LLP  
 STREET: 1185 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: U.S.A.  
 ZIP: 10036

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/728,323A  
 FILING DATE:

## CLASSIFICATION:

CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: White, John P.  
 REGISTRATION NUMBER: 28,678  
 REFERENCE/DOCKET NUMBER: 0575/52268/JPM/MS/SKS  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212-278-0400  
 TELEFAX: 212-391-0525

## INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 3489 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1..3489  
 US-08-728-323A-1

Query Match 4.0%; Score 144; DB 2; Length 3489;  
 Best Local Similarity 45.1%; Pred. No. 3e-26;  
 Matches 534; Conservative 0; Mismatches 650; Indels 0; Gaps 0;

QY 975 GATCAGAGGCTGCTGAGAGCCGTCTCAGAGATGACAGCAGCCAGAGAAAGT 1034  
 DB 1578 GCAGAGCCACAGCAGAGAGCCACAGCAGCAGAGCCACAGCAGAGAGCAGCA 1637  
 QY 1035 GCTGTGACATTTGAAGATPAGAAAGCGGAGAACTTCGAGCCAGGAGTGTGAGCTGGA 1094  
 DB 1638 GCAGGAGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGCA 1697  
 QY 1095 GAAGCCCGCCAGCGCTCTTGAGAGCAGCGCAAGAGCCAGGAGCTTGGCTCAGCT 1154  
 DB 1698 GCGGAGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGCA 1757  
 QY 1155 GAGCGCGCCAGCAGAGAGAAAGCGGAGCCAGCAGCAGAGAGAGCCAGAGCGCA 1214  
 DB 1758 GCGGAGCCACAGCAGCGGAGCCACAGCAGCAGAGATGAGCAGCAGAGATGAGAGCA 1817  
 QY 1215 GCTGAGCTGGAAGAGCAGCTGGAAGAGCAGCGGAGCTGGAAGCGGAGCCAGAGAGCA 1274

DB 1818 GCAGATGACAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGAGATGAGCAGCA 1877  
 QY 1275 GAGAGGAAGAGATCGAGAGCGCGGAGCCGCAAAAAGGAACTGGAAGAGCAGCA 1334  
 DB 1878 GCAGATGACAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGAGATGAGCAGCA 1937  
 QY 1335 ACTTGAATGGAGCGGAACCGGAGCAGGAATCTCTGATCAGAGAAACAAGAGCAGGA 1394  
 DB 1938 GCAGATGACAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGAGATGAGCAGCA 1997  
 QY 1395 GGGCAGCTGCTCTGAAGAGCAAGAGAGAACTCTGAGTTTGAAGTCTTGAA 1454  
 DB 1998 GAGGATGACAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGAGATGAGCAGCA 2057  
 QY 1455 TGACAAAAGCATCAGCTAGAAGGAAAATTCAGATATCAGTGTGCTGCACTGCAACCA 1514  
 DB 2058 GCAGATGACAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGAGATGAGCAGCA 2117  
 QY 1515 GAGCAGAAAATTGAGAGCAGCAAGCAAGTCTAGAGCTTGAAGTTGCTGAAATCACCCA 1574  
 DB 2118 GAGCAGCAGATGAGCAGCAGAGCAGCAGATGAGCAGCAGAGATGAGCAGCAGCA 2177  
 QY 1575 CTTACAGCAGCAGTTGACAGAACTCTACGCAAAATGCTTGAAAGATTTATCCAGGAAACA 1634  
 DB 2178 GAGTGAAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGAGATGAGCAGCA 2237  
 QY 1635 GATACCTAGTGCCAGCTTGAAGCAAGTCACAGCAGCAAGTTTGATGAGACTGCTTCT 1694  
 DB 2238 GCAGCAGCAGCAGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2297  
 QY 1695 TACCTCAAAAAGCCTTGAAGCAAGTGGCCCGCAGCAGCTCCGAGCAGCT 1754  
 DB 2298 GCAGAGATTAGAGGAGCAGCAGCAGAGTTAGAGATCAGAGCAGAGATTAGAGAGCA 2357  
 QY 1755 GAGCAGGTGAGAGAGAGCAGCAGTCAAGCTCAGAGATTTATGTTTCAACCA 1814  
 DB 2358 GAGCAGAGATTAGAGCAGCAGCAGCAGAGTTAGAGCAGCAGCAGCAGCAGCAGCAGCAG 2417  
 QY 1815 GCTGAAGAACTGAGAGAGATACATAGCAAAACAGCACTCAGAAAGCAGAGTCCCTGGA 1874  
 DB 2418 GCAGAGCAGAGATTAGAGAGCAGCAGCAGCAGCAGTTCAGAGCAGCAGCAGCAGCAG 2477  
 QY 1875 GGCAGCGGACTGAAGCAAAAGAGCAGAGAGAGAGAGCCTGAGTTAGAAAGCAAAA 1934  
 DB 2478 GCAGCAGCAGCAGGATTAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2537  
 QY 1935 GAAAGACCTCAGAGAGCAGTTTCAGAAAGGACACAAATGCTGAGCAGATGTCAGCA 1994  
 DB 2538 GCAAGAGCAGAGGTGAGAAAGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2597  
 QY 1995 GAGAGAGCAGCAGCCCGCGGAAACCCACAGAGAGAGCAGACTGAAGAGGGAAGCAG 2054  
 DB 2598 GAGAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2657  
 QY 2055 TGTCAAGAAAGAGGCGGAAAGAGAGAGCCAGCCGGAATGCAAGCAAGCAGACTCG 2114  
 DB 2658 GCAGAAAGAGCAGAGATTAGAGAGAGTGAAGAGCAGAAAGAGAGAGATTAGAGAGGT 2717  
 QY 2115 GCTTTTCATCCGATCAGAGAGCAGCTTAAGCTGCGCACCCAGG 2158  
 DB 2718 GAAAGAGCAGAGCAGCAGAGATTAGAGAGGTGGAAGAGCAGG 2761

## RESULT 7

US-09-298-568-1  
 ; Sequence 1, Application US/09298568

## GENERAL INFORMATION:

APPLICANT: Kieff, Elliott D.  
 APPLICANT: Ballestas, Mary E.

TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO  
 ; TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE

FILE REFERENCE: 16412-10001R  
 CURRENT APPLICATION NUMBER: US/09/298,568  
 CURRENT FILING DATE: 1999-04-21  
 EARLIER APPLICATION NUMBER: US 60/109,422  
 EARLIER FILING DATE: 1998-11-19  
 NUMBER OF SEQ ID NOS: 3  
 SOFTWARE: Patent In Ver. 2.0  
 SEQ ID NO: 1  
 LENGTH: 3489  
 TYPE: DNA  
 ORGANISM: Kaposi's sarcoma-associated herpesvirus  
 US-09-298-568-1

Query Match 4.0%; Score 144; DB 4; Length 3489;  
 Best Local Similarity 45.1%; Pred. No. 3e-26;  
 Matches 534; Conservative 0; Mismatches 650; Indels 0; Gaps 0;

QY 975 GATTCAGAGGCTGCTGAGAGAGCCGTCGTCAGAGATGAGCCAGCCAGAGAACT 1034  
 DB 1578 GCAGGAGCCACAGCAGCAGAGAGCCACAGCAGGAGCCACAGCAGCAGAGACACAGCA 1637  
 QY 1035 GCCTGTGACATTGAGATTAAGAGCGGAGAACTTGAAGCGAGAGAGTGTGAGCTTGA 1094  
 DB 1638 GCAGGAGCCACAGCAGCAGAGAGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGCA 1697  
 QY 1095 GAAGCCGCGCAAGCGCTCTTGGAGCAGCAGCGCAAGAGCAGAGCGCTTGAAGCT 1154  
 DB 1698 GCGGAGCCACAGCAGCAGAGAGCCACAGCAGCGGAGCCACAGCAGCGGAGCCACAGCA 1757  
 QY 1155 GAGAGCGCGCCAGCAGAGAGAGAGAGAGCGGAGCCAGCAGCAGAGAGCGCAAGCGCA 1214  
 DB 1758 GCGGAGCCACAGCAGCAGAGAGAGCCACAGCAGCGAGATGAGCAGCAGCAGAGATGAGCAGCA 1817  
 QY 1215 GCTGAGAGCTGAGAGAGAGCTGAGAGAGCAGCGGAGCTTGAAGCGCAGCAGAGAGCA 1274  
 DB 1818 GCAGGATGACAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCAGCAGAGATGAGCAGCA 1877  
 QY 1275 GAGGAGAGAGAGATGAGAGAGCGCCAGCGCCAGAGAGAGAGAGAGAGAGAGAGAGCA 1334  
 DB 1878 GCAGGATGACAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCAGCAGAGATGAGCAGCA 1937  
 QY 1335 ACTTGAATGGAGCA 1394  
 DB 1938 GCAGGATGACAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCAGAGAGATGAGCAGCA 1997  
 QY 1395 GAGGAGCA 1454  
 DB 1998 GCAGGATGACAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCAGAGAGATGAGCAGCA 2057  
 QY 1455 TGACAAAGAGATGAGCAGCAGCA 1514  
 DB 2058 GCAGGATGACAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCAGAGAGATGAGCAGCA 2117  
 QY 1515 GAGCAGCA 1574  
 DB 2118 GCAGCAGAGAGATGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCA 2177  
 QY 1575 CTTACAGCAGAGATGAGAGAGAGATGAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGCA 1634  
 DB 2178 GCAGCAGCAGCAGCAGCA 2237  
 QY 1635 GATACAGCAGCA 1694  
 DB 2238 GCAGCAGCAGCAGCA 2297  
 QY 1695 TACCTCAAGCA 1754  
 DB 2298 GCAGCA 2357  
 QY 1755 GAGCA 1814  
 DB 2358 GAGCA 2417

QY 1815 GCTGAGCA 1874  
 DB 2418 GCAGGAGCA 2477  
 QY 1875 GCGAGCCGAGCA 1934  
 DB 2478 GCAGCAGCA 2537  
 QY 1935 GGAAGAGCTGAGCA 1994  
 DB 2538 GCAAGAGCAGCA 2597  
 QY 1995 GAGCA 2054  
 DB 2598 GAGCA 2657  
 QY 2055 TGTCAGCA 2114  
 DB 2658 GCAGCA 2717  
 QY 2115 GCTTTTCATCCGATCAGCA 2158  
 DB 2718 GGAAGCA 2761

RESULT 8  
 US-08-770-379-20/c  
 Sequence 20, Application US/08770379  
 Patent No. 5849564

GENERAL INFORMATION:  
 APPLICANT: Chang, Yuan  
 APPLICANT: Bohenzky, Roy A.  
 APPLICANT: Russo, James J.  
 APPLICANT: Edelman, Isidore S.  
 APPLICANT: Moore, Patrick S.  
 TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED  
 TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF  
 NUMBER OF SEQUENCES: 20  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: Cooper & Dunham LLP  
 STREET: 1185 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: U.S.A.  
 ZIP: 10036

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/770,379  
 FILING DATE:

CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: White, John P.  
 REGISTRATION NUMBER: 28,678  
 REFERENCE/DOCKET NUMBER: 52342  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 278-0400  
 TELEFAX: (212) 391-0525  
 INFORMATION FOR SEQ ID NO: 20:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3207 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-08-770-379-20

Query Match 4.0%; Score 144; DB 2; Length 3207;  
 Best Local Similarity 45.1%; Pred. No. 9.3e-26;  
 Matches 534; Conservative 0; Mismatches 650; Indels 0; Gaps 0;

QY 975 GGATCAGAGGCTGCTTGAAGAGCCGTCGTCAGAGATGAGCAGACCCAGAGAAACT 1034  
 Db 20419 GCAGAGGCCACAGCAGAGAGGCCACAGCAGCAGAGGCCACAGCAGAGGACAGCA 20360  
 QY 1035 GCTGTGACACTTTGAAGATTAAGAAAGGGGAACTTTCAGCCGAGCAGTGTGAGCTGGA 1094  
 Db 20359 GCAGAGGCCACAGCAGAGGAGGCCACAGCAGCAGGCCACAGCAGAGGACAGCA 20300  
 QY 1095 GAAGCGCGCCGCAAGCGCTCTTGAGCAGCAGCCAAAGAGCAGAGCGGTGGCTGAGCT 1154  
 Db 20299 GCGGAGGCCACAGCAGAGGAGGCCACAGCAGCAGGCCACAGCAGAGGAGCCACAGCA 20240  
 QY 1155 GAGCGCGCCGCAAGCAGAGAGAAAGCGGAGCCGCAAGCAGCAGAGGCCAAGCGGCA 1214  
 Db 20239 GCGGAGGCCACAGCAGAGGAGGCCACAGCAGCAGGATGAGCAGCAGCAGAGATGAGCAGCA 20180  
 QY 1215 GCTTGAGCTGAGAGAGCAGTGGAGAAAGCAGCGGAGCTTGAGCGGCGCAGCAGAGAGA 1274  
 Db 20179 GCAGGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCA 20120  
 QY 1275 GAGAGAGAAAGAGATGAGAGGCGCGAGCGCCGCAAAACGGAACTGGAAAGCAGCAGCA 1334  
 Db 20119 GCAGGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCA 20060  
 QY 1335 ACTTGATGGGAAACGGAACCGAGACAGGAACCTCTGAATCAGAGGAAACAGAGCAGCA 1394  
 Db 20059 GCAGGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCA 20000  
 QY 1395 GGGCACCCTGCTCTTGAAGCAGAGAAAGACTCTGAGTTTGAAGTGAAGCTCTGAA 1454  
 Db 19999 GCAGGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCA 19940  
 QY 1455 TGCAAAAAGCATCAGCTTGAAGAAACTTTCAGGATTCAGTGTGCACTGCAACCCA 1514  
 Db 19939 GCAGGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCA 19880  
 QY 1515 GAGCAGAAATTTGAGAGCAGCAAAAGCTTAGAGACTTAAGATTTGTAATTCACCCA 1574  
 Db 19879 GAGCAGCAGCAGATGAGCAGCAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCA 19820  
 QY 1575 CTTCACAGCAGCAGTTCAGAGAACTTCAGCAAAATGCTTGAAGAATTATTCAGAGAAACA 1634  
 Db 19819 GATGAGCAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCA 19760  
 QY 1635 GATATCTAGTGAACAGTTAAACAGTCCAGCAGCAAGTTTGATAGAGCTCGCTTCT 1694  
 Db 19759 GCAGCAGCAGCAGATGAAACAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 19700  
 QY 1695 TACCTCAAAAAGCCTTGGAAACAAAGAGCTGCGCCGAGCAGCAGCTCCGGAGCAGCT 1754  
 Db 19699 GCAGGAGTTAGAGAGCAGAGCAGAGGATTAAGAGATCAGAGCAGCAGAGTTAGAGAGCA 19640  
 QY 1755 GAGCAGGTGAGAGAGAGCAAGCTCAAGCTCAAGCTCAGAGATTTGATTTTCAACAACA 1814  
 Db 19639 GAGCAGAGATTTAGAGAGCAGAGCAGAGGATTAAGAGCAGCAGCAGAGATTTAGAGA 19580  
 QY 1815 GCTGAAGAACTAGAGAGATTAATAGCAAAACAGCAATTCAGAGCAGAGAGTCTCTGGA 1874  
 Db 19579 GCAGGAGCAGAGATTTAGAGAGCAGAGCAGAGGATTAAGAGCAGCAGAGATTTAGA 19520  
 QY 1875 GAGAGCGCCGCTGAAGCAGAAAGCAGAGAGAGAGAGCTTGAATTGAGAAAGCAAA 1934  
 Db 19519 GAGCAGAGCAGCAGATTTAGAGAGCAGAGCAGAGGATTAAGAGCAGCAGAGTGAAGA 19460  
 QY 1935 GGAAGACGCTCAGAGAGCAGATTTAGAGAAAGAGCAAGCAATGCTTGAGCATGTGACGA 1994  
 Db 19459 GCAAGAGCAGAGAGTGAAGAGCAAGAGCAGAGCAGAGAAAGCAAGAAATTAGAGAGGT 19400  
 QY 1995 GAGAGCAGCAGCAGCGCCCGGAAACCCACAGAGAGCAGCAGACTGAAGAGAAACAG 2054  
 Db 19399 GAGAGAGCAAGCAGAGAGCAGAGAGAGAGAGAGAGAGAGATTTAGAGAGGTGGAAGA 19340

QY 2055 TGTCAAGAAAGAGAGCGCGAGAGAGAGCCCAAGCCGAAATGCAAGACAGAGACTG 2114  
 Db 19339 GCAGAGAGCAGCAGAGTTAGAGAGGTGGAAGAGCAGAGAGCAGAGAGTTAGAGAGGT 19280  
 QY 2115 GCTTTCCATCCGATCAGAGGCCAGCTTAAGCTGGCCACCAGG 2158  
 Db 19279 GGAAGAGCAGAGAGCAGCAGAGTTAGAGAGGTGGAAGAGCAGG 19236

RESULT 9  
 US-08-757-669A-20/C  
 ; Sequence 20, Application US/08757669A  
 ; Patent No. 6183751  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Chang, Yuan  
 ; APPLICANT: Bohenzky, Roy A.  
 ; APPLICANT: Russo, James J.  
 ; APPLICANT: Edelman, Isidore S.  
 ; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS  
 ; NUMBER OF SEQUENCES: 20  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Cooper & Dunham LLP  
 ; STREET: 1185 Avenue of the Americas  
 ; CITY: New York  
 ; STATE: New York  
 ; COUNTRY: U.S.A.  
 ; ZIP: 10036  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/757,669A  
 ; FILING DATE:  
 ; CLASSIFICATION: 424  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: White, John P.  
 ; REGISTRATION NUMBER: 28,678  
 ; REFERENCE/DOCKET NUMBER: 45185-F  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (212) 278-0400  
 ; TELEFAX: (212) 391-0525  
 ; INFORMATION FOR SEQ ID NO: 20:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 32207 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; US-08-757-669A-20

Query Match 4.0%; Score 144; DB 4; Length 32207;  
 Best Local Similarity 45.1%; Pred. No. 9,3e-26;  
 Matches 534; Conservative 0; Mismatches 650; Indels 0; Gaps 0;

QY 975 GGATCAGAGGCTGCTTGAAGAGCCGTCGTCAGAGATGAGCAGACCCAGAGAAACT 1034  
 Db 20419 GCAGAGGCCACAGCAGAGAGGCCACAGCAGCAGAGGCCACAGCAGAGGACAGCA 20360  
 QY 1035 GCTGTGACACTTTGAAGATTAAGAAAGGGGAACTTTCAGCCGAGCAGTGTGAGCTGGA 1094  
 Db 20359 GCAGAGGCCACAGCAGAGGAGGCCACAGCAGCAGGCCACAGCAGAGGACAGCA 20300  
 QY 1095 GAAGCGCGCCGCAAGCGCTCTTGAGCAGCAGCCAAAGAGCAGAGCGGTGGCTGAGCT 1154  
 Db 20299 GCGGAGGCCACAGCAGAGGAGGCCACAGCAGCAGGATGAGCAGCAGCAGAGATGAGCAGCA 20240  
 QY 1155 GAGCGCGCCGCAAGCAGAGAGAAAGCGGAGCCGCAAGCAGCAGAGGCCAAGCGGCA 1214  
 Db 20239 GCGGAGGCCACAGCAGAGGAGGCCACAGCAGCAGGATGAGCAGCAGCAGAGATGAGCAGCA 20180





QY 1755 GAGCGAGTGTGAGAGAGACCAAGTCAAACTGCAGAGATTGATGTTTCAACACCA 1814  
 DB 19639 GGAGCAGGAGTTAGAGAGCAGAGCAGAGTTAGAGACAGACAGAGATTAGAGA 19580  
 QY 1815 GCTGAAGAACTGAGAGATACATAGCAACGCAACTCCAGACGAGGCTCCCTGGA 1874  
 DB 19579 GCAGGAGCAGAGCTTAAAGAGAGAGAGAGATTAGAGACAGAGAGAGATTAGA 19520  
 QY 1875 GAGAGCGGAGTGAAG 1934  
 DB 19519 GGAGCAGGAGAGAGAGATTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 19460  
 QY 1935 GGAAGAGCTTCAG 1994  
 DB 19459 GCAAGAGCAG 19400  
 QY 1995 GAGAGGAG 2054  
 DB 19399 GAG 19340  
 QY 2055 TGTCAG 2114  
 DB 19339 GCAG 19280  
 QY 2115 GCTTTTCATCCGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2158  
 DB 19279 GGAAGAGCAG 19236

## RESULT 11

US-09-404-879A-4/c  
 ; Sequence 4, Application US/09404879A  
 ; Patent No. 6468546  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Mitcham, Jennifer L.  
 ; APPLICANT: King, Gordon E.  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
 ; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER  
 ; FILE REFERENCE: 210121.462C2  
 ; CURRENT APPLICATION NUMBER: US/09/404, 879A  
 ; CURRENT FILING DATE: 1999-09-24  
 ; NUMBER OF SEQ ID NOS: 393  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 4  
 ; LENGTH: 531  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapien  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (1)-(531)  
 ; OTHER INFORMATION: n = A,T,C or G  
 US-09-404-879A-4

## Query Match

Best Local Similarity 53.5%; Pred. No. 4,8e-21;  
 Matches 250; Conservative 0; Mismatches 217; Indels 0; Gaps 0;

QY 1384 AAGGAGCAGAGAGGCGACCGTGTCTGGAAGCAAGAGAGAGAGAGAGAGAGAGAGAG 1443  
 DB 531 AGGAGACAAAG 472  
 QY 1444 GAGGCTGTAATGACAAAG 1503  
 DB 471 GAGGCTGTAATGAG 412  
 QY 1504 CTGGCAACCCAG 1563  
 DB 411 AAGCAAAACAG 352  
 QY 1564 GAAATCCCACTTACGAG 1623

DB 351 GAAATCAACCACTTCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 292  
 QY 1624 CCAGAGAAACAGATCTAGTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1683  
 DB 291 CCTGAG 232  
 QY 1684 GACTGCTTTTACCTTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1743  
 DB 231 TCAGGAGATCAGTTTACCTTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 172  
 QY 1744 CGGAG 1803  
 DB 171 AAGAGACATTAAG 112  
 QY 1804 TTCAACCAACAG 1850  
 DB 111 TTTAATCACTGAG 65

## RESULT 12

US-08-056-200-93  
 ; Sequence 93, Application US/08056200  
 ; Patent No. 5615500  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Steinert, Peter M.  
 ; APPLICANT: Lee, Seung-Chul  
 ; APPLICANT: Kim, In-Gyu  
 ; APPLICANT: Chung, Soo-Il  
 ; APPLICANT: Park, Sang-Chul  
 ; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
 ; TITLE OF INVENTION: Methods of Using Same  
 ; NUMBER OF SEQUENCES: 117  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Knobbe, Martens, Olson & Bear  
 ; STREET: 620 Newport Center Drive, Sixteenth Floor  
 ; CITY: Newport Beach  
 ; STATE: CA  
 ; COUNTRY: U.S.A.  
 ; ZIP: 92660  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/056,200  
 ; FILING DATE: 30-APR-1993  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Fedrick, Michael F.  
 ; REGISTRATION NUMBER: 36,799  
 ; REFERENCE/DOCKET NUMBER: NIH054.001A  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (714) 760-0404  
 ; TELEFAX: (714) 760-9502  
 ; INFORMATION FOR SEQ ID NO: 93:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 9551 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: cDNA  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 1507..1644  
 ; FEATURE:  
 ; NAME/KEY: intron  
 ; LOCATION: 1645..2511  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 2512..8070

US-08-056-200-93

Query Match	3.3%;	Score 119.6;	DB 1;	Length 9551;
Best Local Similarity	44.1%;	Pred. No. 6.5e-20;		
Matches 500;	Conservative	0;	Mismatches 634;	Indels 0;
				Gaps 0

OY	975	GGATCAGAGGCGTGGCTTACAGGAGCGCTGTGTACAGAGTATGACGACGACGACGAGAAAGAACT	1014
Db	3324	GCAAGAGGAGAGGCGCGAGCGACGAGGAGAGCGCGGACGACGAGAGGAGAGAGCGCGACGA	3383
OY	1035	GCTGTGACATTTTGAAATTAAGAACGGGAGAACTTTCAGACGACGACGCTGTGAGCTTGA	1094
Db	3384	GCAGCTGAGGCGCGACGACGAGAGGAGAGGAGCGCGACGACGACGCTGTGAGCGCGGACGAGAGGA	3443
OY	1095	GAAGCGCGCCCAACGCGCTTTTGGAGCGACGACCCCAAGAGCAGAGCGGTTGTGCTACT	1154
Db	3444	GGAAGCGCGCGACGACGACCGCTGAGCGCGCGACGAGAGGAGAGAGCGCGGACGACGACT	3503
OY	1155	GGAGCGCGCCGACGACGAGAGAGAGAAAGACCGGAGCGCCACAGAGCAGAGAGCCCAAGCGCA	1214
Db	3504	GAGCGCGCAGACGAGAGGAGAGAGCGCGACGACGCTGAGCGCGGACGACGACGCTGAG	3563
OY	1215	GCTGGAGCTTGAGAAAGCAGCTTGAGAAAGCAGCGGAGCTGGAGCGGACGACGAGAGAGGA	1274
Db	3564	GCGCGACGACGACGCTGAGAGCGCGACGACGACGCTGAGAGCGCGACGACGACGCTGAGCGCGA	3623
OY	1275	GAGAGGAAAGAGATTCGAGAGCGCGACGCGCCGAAAAACGGAACTTGGAAAGCAGCGCA	1334
Db	3624	GCACGAGCTTGAGGCGCGACGACGCTGAGCGCGACGACGCTGAGCGCGGACGACGA	3683
OY	1335	ACTTAATGGGAAACGGAAACCGGAAACCGGAACTCTGAAATCAGAGAACAAAGAGACGA	1394
Db	3684	GCTGAGGCGCGACGACGACGCTGAGCGCCGACGAGAGAGAGAGAGGCAAGACGAGACGA	3743
OY	1395	GGGCACCGTGTCTTGAAGCGAGAGAGAAAGACTCTTGGAGTTTGAATTGAAGCTTCTAA	1454
Db	3744	CGAGCAGAGAGAGGCGCGAGCGCGCTGAAAGCGCGACGAGAGAGAGCGCGAATTGCT	3803
OY	1455	TGACAAAAAGCATCAGCTAGAGAGAAAAATTACAGATTCAGGTGTGCACTGGCAACCA	1514
Db	3804	GAGGCGCAGAGAGAGACGAGAGAGGACGACGACGAGAGAGCGCCAAACGACGACTGAACG	3863
OY	1515	GAGCGAAGAAATTGAGACGACGAACAAGTCTAGAGACTTAAAGATTCTGAATATCACCA	1574
Db	3864	CGACCGAGAGAGAGGAGCGCGCAACCTTGTGTAAGCTTGAAGAGAGAGAGAGCGCGGA	3923
OY	1575	CTTACGAGGACGTTTGCAGGAATCTCAGCAAAATGTCTTGGAAGACTTAATTCAGAGAAACA	1634
Db	3924	GCAGCAGAGAGGCGCGACGACCACTTAAGCGGAGCAAGAGAGAGCGCGGAGACGCG	3983
OY	1635	GATTACTCAGTGAACGATTAAAAACAAGTCCAGCAAAACAGTTTGCATAGACTCGCTTCT	1694
Db	3984	GCTTGAAGCGCAGGAGGAGGAAGAGAGGCTCCAGCAGCGGTTGAGAGCGAACAACAT	4043
OY	1695	TACCTTCAAAAGAGCTTTGGAAGCAAAAGAGAGCTGGCCCGGACGACACTCCGGAGCACCT	1754
Db	4044	AAGACGCGACGAGGAGAGAGAGGCTTCAGCAGCTTCTGAAACGCGAGAGAGAGAAAGAGCT	4103
OY	1815	GTTGAAGAACTGAGAGAGATTAATACCAAAACGCAACTCTCAGAACGACAGAGTCCCTCGA	1874
Db	4164	GCTTGAAGCGCAGAGAGAGAGAGCGCCACGACGCGCTGAAACCGCGACAGAAAGAGAGCT	4223
OY	1875	GGCAGCGGACCTGAAAGCAGAAAGAGCAGAGAGAGAGAGCGCTTGAGTTGAGAAACAAA	1934
Db	4224	CGAGCAGCGACTGAAGCGCGAGAGGTTGAGAGACTTGACACAGAGAGAGCGCGACGA	4283
OY	1935	GGAAACCGCTCAGAGACGAGTTCCAGGAAAGGACCAAGCAATGGCTTGGACATCTGTCCACA	1994
Db	4284	GCGGCTGAAGCGCGAGAGCGCGAGAGAAAGAGCGCGCACAGACTCTCTTAAAGACGAGAGA	4343

Oy	1995	GGAGACGACACACACGCCCCCGAAACCCACAGAGAGACACAACTGAAGAGAGGAAAGACG	2054
Db	4334	GCNCGAGAGAGGCGCCAGAGCACTGAGGCGCGAGCAGCAGAAAGCGCGAGCAGCG	4403
Oy	2055	TGTCAGAAAGAGAGCGGGAAGAGAGACCAAGCCGGAATGCAAGACAAGCA	2108
Db	4404	GCTTAAGCGGAGAGAGAGAAAGAGGCTTCAGACGACGCGCTGAAGCCGCAACA	4457

RESULT 13  
US-08-800-644-93

GENERAL INFORMATION:

APPLICANT: Kim, In-Gyu

TITLE OF INVENTION: Trichohyalin

CORRESPONDENCE ADDRESS:

CITY: Newport Beach

ZIP: 92660

COMPUTER: IBM PC compatible

CURRENT APPLICATION DATA:

CLASSIFICATION: 424  
FILING DATE: 14-FEB-1991

APPLICATION NUMBER: US 08/  
 FILING DATE: 30-APR-1993

NAME: Fedrick, Michael F.  
REGISTRATION NUMBER: 36 799

TELECOMMUNICATION INFORMATION  
BRIEFING (714) 760-0404

INFORMATION FOR SEQ ID NO: 93:

TYPE: nucleic acid

MOLECULE TYPE: CDNA

FEATURE:

FEATURE:

FEATURE:

8-800-644-93

Local Similarity 44.1%;

100

Oy	1995	GGAGACGACACACACGCCCCCGGAAACCCACAGAGAGACACAACTGAAGAGAGGAAAGACG	2054
Db	4334	GCNCGAGGAGGCGCCAGAGCACTGAGGCGCGAGCAGCAGGAAAGCGCGAGCAGCG	4403
Oy	2055	TGTTCAGAAAGAGAGCGGGAAGAGAGACCAAGCCGGAATTCAGAGCAAGCA	2108
Db	4404	GCTTAAGCGGAGAGAGAGAAAGAGCGCTCGAGCAGCGCGCTGAAGCCGAGCA	4457

Query Match	3.3%	Score 119.6	DB 2	Length 9551
Best Local Similarity	44.1%	Pred. No. 6.5e-20		
Matches 500	0	Mismatches 634	Indels 0	Gaps 0



```

NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/050,863
FILING DATE: 30-MAR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 949-8711
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2580 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
US-09-050-863-2

Query Match      3.3%; Score 119.2; DB 3; Length 2580;
Best Local Similarity 53.4%; Pred. No. 4.2e-20;
Matches 250; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

QY 973 GTGGATCAGAGGCTGCTAGAGAGCCGCTGTCAGAGATGACGACCCAGAGAAAGAA 1032
DB 898 GAGAGGGGCGAGAGCAGAGAGGGCGAGAGGGCGAGAGCAGAGAGGGGCGAGAG 957

QY 1033 CTGCTGTGACATTGTAAGATTAAGACCGGGAGAACTTCGAGCGAGGCACTGTGAGCTG 1092
DB 958 GGGCAGAGGGGCGAGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1017

QY 1093 GAGAAGCGCCGCAAGCGCTCTTGAGCAGCAGCGCAAAAGCAGAGCGGTTGCTCAG 1152
DB 1018 GGGCAGAGCAGAGAGGGCGAGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1077

QY 1153 CTGAGCGCGCCGAGCAGAGAGAGAAAGAGCGGAGCGCCAGAGCAGAGAGCCAGCGG 1212
DB 1078 CAGAGGGGCGAGAGCAGAGAGGGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1137

QY 1213 CAGCTGAGCTGAGAGCAGACTGAGAAAGCAGCGGAGCTGAGCGGAGCGAGAGAGAG 1272
DB 1138 GGGCAGAGCAGAGAGGGCGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1197

QY 1273 GAGAGAGAGAAAGAGATGAGAGGCGCGAGGCGCAAAACGGGAATCTGAAAAGGCGAGA 1332
DB 1198 GAGAGGGGCGAGAGGCGAGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1257

QY 1333 CAATCTGAATGGGAACGGAACCGGAGACAGGAATCTCTGAATCAGAGGAACAAAGAGCAG 1392
DB 1258 CAGAGGGGCGAGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1317

QY 1393 GAGGAGCACTGTGCTCTGAAAGGCAAGAGAGAACTCTGAGATTGAG 1440
DB 1318 GAGCAGAGAGAGGGCGAGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1365
```

Search completed: March 14, 2003, 11:19:06  
Job time : 358.517 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2003, 01:42:46 ; Search time 665.293 Seconds  
(without alignments)  
3842.205 Million cell updates/sec

Title: US-09-674-237A-2  
Perfect score: 3642  
Sequence: 1 atggctcagttccacacaccc.....cggaccaccagcagcatga 3642

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 501302 seqs, 350932545 residues  
Total number of hits satisfying chosen parameters: 1002604

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: Published Applications\_NA.\*  
2: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	477.4	13.1	2017	10	US-09-884-441-72 Sequence 72, Appl
2	332.4	9.1	2873	10	US-09-879-957-193 Sequence 125, App
3	331.4	9.1	4210	9	US-09-764-868-125 Sequence 125, App
4	258.9	7.1	747	10	US-09-879-957-39 Sequence 39, Appl
5	198.8	5.5	270	10	US-09-864-761-17127 Sequence 17127, A
6	198.8	5.5	286	10	US-09-864-761-26948 Sequence 26948, A
7	198.8	5.5	297	10	US-09-864-761-30453 Sequence 30453, A
8	198.6	5.5	301	10	US-09-864-761-17146 Sequence 17146, A
9	197.4	5.4	263	10	US-09-864-761-17644 Sequence 17644, A
10	192	5.3	480	10	US-09-864-761-10314 Sequence 10314, A
11	180.8	5.0	486	10	US-09-864-761-333 Sequence 333, App
12	180.8	5.0	487	10	US-09-864-761-864 Sequence 864, App
13	163	4.5	475	10	US-09-864-761-311 Sequence 311, App
14	163	4.5	475	10	US-09-864-761-13884 Sequence 13884, A
15	146.4	4.0	180	10	US-09-864-761-17125 Sequence 17125, A
16	144.2	4.0	292	10	US-09-864-761-20261 Sequence 20261, A
17	144.2	4.0	304	10	US-09-864-761-21373 Sequence 21373, A
18	144.2	4.0	310	10	US-09-864-761-19751 Sequence 19751, A
19	144.2	4.0	310	10	US-09-864-761-19759 Sequence 19759, A

20	144	4.0	480	10	US-09-884-441-60 Sequence 60, Appl
21	144	4.0	531	10	US-09-884-441-5 Sequence 5, Appl
22	140.4	3.9	230	10	US-09-864-761-17643 Sequence 17643, A
23	140.4	3.9	247	10	US-09-864-761-30501 Sequence 30501, A
24	127	3.5	419	10	US-09-864-761-309 Sequence 309, App
25	125.2	3.4	484	10	US-09-864-761-863 Sequence 863, App
26	121.6	3.3	531	10	US-09-884-441-4 Sequence 4, Appl
27	118.4	3.3	2108	10	US-09-864-761-325 Sequence 225, App
28	117	3.2	466	10	US-09-864-761-3491 Sequence 3491, App
29	117	3.2	466	10	US-09-864-761-4631 Sequence 4631, App
30	113	3.1	477	10	US-09-864-761-11936 Sequence 11936, A
31	107.8	3.0	472	10	US-09-864-761-310 Sequence 310, App
32	107.8	3.0	472	10	US-09-864-761-14159 Sequence 14159, A
33	105.8	2.9	198	10	US-09-864-761-17126 Sequence 17126, A
34	105.8	2.9	198	10	US-09-864-761-30714 Sequence 30714, A
35	97.6	2.7	458	10	US-09-864-761-13834 Sequence 13834, A
36	96	2.6	463	10	US-09-864-761-2978 Sequence 2978, App
37	96	2.6	465	10	US-09-864-761-2970 Sequence 2970, App
38	95.6	2.6	206	10	US-09-864-761-30400 Sequence 30400, A
39	80	2.2	381	10	US-09-884-441-61 Sequence 61, Appl
40	79.2	2.2	474	10	US-09-864-761-11284 Sequence 11284, A
41	79.2	2.2	720	9	US-10-101-487-74 Sequence 74, Appl
42	79.2	2.2	720	9	US-10-101-487-76 Sequence 76, Appl
43	78.4	2.2	3348	10	US-09-312-762A-2 Sequence 2, Appl
44	78.4	2.2	14707	10	US-09-312-762A-3 Sequence 3, Appl
45	78.2	2.1	267	10	US-09-864-761-27984 Sequence 27984, A

#### ALIGNMENTS

RESULT 1									
US-09-884-441-72									
Sequence 72, Application US/09884441									
Patent No. US20020119158A1									
GENERAL INFORMATION: Homo sapien									
APPLICANT: Algate, Paul A.									
TITLE OF INVENTION: CARTEK, DARRICK									
TITLE OF INVENTION: DIAGNOSIS AND METHODS FOR THE THERAPY AND									
FILE REFERENCE: 210121.462C7									
CURRENT APPLICATION NUMBER: US/09/884,441									
CURRENT FILING DATE: 2001-06-18									
NUMBER OF SEQ ID NOS: 489									
SOFTWARE: FastSeq for Windows Version 3.0									
SEQ ID NO 72									
LENGTH: 2017									
TYPE: DNA									
ORGANISM: Homo sapien									
US-09-884-441-72									
Query Match									
Best Local Similarity 13.1%; Score 477.4; DB 10; Length 2017;									
Best Local Similarity 55.5%; Pred. No. 2.3e-112;									
Matches 111; Conservative 0; Mismatches 796; Indels 96; Gaps 6;									
OY	1	ATGGCTCAGTTTCCACACCTTTGCGTGTAGCTGTGATGTCGGCCATACGTGAG	60						
DB	42	ATGGCTCAGTTTCCACACCTTTGCGTGTAGCTGTGATGTCGGCCATACGTGAG	101						
OY	61	GAAGGCGCAACATGACACGACCTTCTAGCCTGAAGCCGATGAGGATTTATTA	120						
DB	102	GAAGGCGCAACATGACACGACCTTCTAGCCTGAAGCCGATGAGGATTTATTA	161						
OY	121	GGATGCAAGGAGAGGAGCTTTTCCATCTGGGTACTCTAGCCTGCTTATGACAA	180						
DB	162	GGATGCAAGGAGGAGCTTTTCTCTACAGTCTGCGGCCCGGTTTATGCTGAA	221						
OY	181	ATATGGCGCTAGCGGACATGATATGATGAGGATGATGATGATGATGATGATG	240						
DB	222	ATATGGCGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT	281						
OY	241	GCCATGAAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATG	300						



Db 282 GCTATGAACTCATCAAGTTAAAGTTSCAGGGCCAAAGCTGCTGAGTCTCCCTCT 341  
 Qy 301 GTATGAAAGCAAGCAAGG---GCTATTTCAAGTGCACAGACATTTGGTATAGAGG 357  
 Db 342 ATATGAAAGCAAGCCCTATGTTCTCTCACTATCTCTGCTGTTGGAGAGGAGC 401  
 Qy 358 ATTGCTAGCATG-----CCACCACTCAAGCTGTGCTCTGTCCTGCA----- 400  
 Db 402 ATGCGCAATCTGTCATTCATTCAGCTCATGCTCAGTTGACCTTATAGCAACCTT 461  
 Qy 401 -----TGGGCTCATTCAGCTGTGGAATGTCTCCACCTTATAGTCT 444  
 Db 462 TCTTCTGCTACTGAGGACCAATATCTCCCTAAATGATGCTGCTCCCTAGTGCCT 521  
 Qy 445 TCTGCTCCCTCAGAGAGAGTGCCTCCCTGCTAACGCGGCTCCTCCCTCATACAGCT 504  
 Db 522 TCTGTTAGTACATCTCTATTAACAAATGGAATGCCAGTCTCATTTACGCTTTATCATT 581  
 Qy 505 CTGCTGCTGTTGGCATCTCTGAGCCACATGCGCAAGATTTCTTCTTACAGAGATCT 564  
 Db 582 CTTATTTCTTCTTCAACATTGCTCATGCAATCTTACAGCTGATGATGGAGGATTT 641  
 Qy 565 GGTCCAGGTGACAAATTAACATAAGTTACAGAGGCAATCATTCATGTCGCGAGC 624  
 Db 642 GGTGCTGATGATCCAGAAAGGCCAGTCTGTGATTTAGGATTTAGTACTGACT 701  
 Qy 625 GCCCTCTCAG-----CAGCAGATGGGCTGTG 651  
 Db 702 TCTCTCAACTGTTCCCTCAGAGGACCTACCTAAGACAGGACCTCAGAGTGGGAGTT 761  
 Qy 652 CCTAGTATCAAGGCTGAATTAACAGGAGTTATTAACAGCCACGCAAACTATGAGT 711  
 Db 762 CTTAGGCTTCAAGATTAAGATATGCGCAAAATTTAATAGTCTAGCAAAAGGATGAGC 821  
 Qy 712 GGAACCTTAACAGGTCCCGAGGCAAGAACTATTCTCATGCAATCAATTTACCCAGCT 771  
 Db 822 GGAATCTCTCAGGTTTCAAGCTGAGAAATGCCCTTCTTCAAGTCAATCTCTCTCAAACT 881  
 Qy 772 CAGCTGCTTCAATATGAGATCTTTCTGCAATGATCAAGATGAGAACTCATCTCAGAA 831  
 Db 882 CAGTACTACTATTTGAGCTGCTGACATGATGATGATGATGATGATGATGATGATGAT 941  
 Qy 832 GAATTTATCCAGTATGACCTAATGATGATGATGATGATGATGATGATGATGATGATGAT 891  
 Db 942 GAATTTATCTGAGGATGACCTCATGACATGCGCAAAAGCTGACACCACTACCACTG 1001  
 Qy 892 GTCCCTGCTCAGAAATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 951  
 Db 1002 ACGTTGCTCCGAGCTGTGCTCCCTCAATCTTTAGAGGGGAAAGCAAG----- 1050  
 Qy 952 TCGCTCATAGCTCTTCT 1011  
 Db 1051 -----TTGATTTCTGTTAATGAACTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1097  
 Qy 1012 GAGCAGCAGCAGAGAGAACTGCTGTGATCTTTGAAGATAAGAGCGGAGAACTTC 1071  
 Db 1098 GA---AAGCTCAGAGAACTGCTGATCTTTTGAAGAGCAAGAGAAAGCACTAT 1154  
 Qy 1072 GAGCAGGCACTGTGAGCTGAGAGAGCGCGCAAGCGCTCTTGAAGCAGAGCGCAAA 1131  
 Db 1155 GAAAGAGAGAAATGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1214  
 Qy 1132 GAGCAG 1191  
 Db 1215 GAGGCTGAAGCAAG 1274  
 Qy 1192 CAGAGCAG 1251  
 Db 1275 CAAAG 1334  
 Qy 1252 CTGAG 1311  
 Db 1335 CTGAG 1394

Qy 1312 CGGGAAGTGAAG 1371  
 Db 1395 CAGAGAGCTTGAG 1454  
 Qy 1372 AATCAG 1431  
 Db 1455 AGTCAG 1514  
 Qy 1432 GAGTTGAGTTAG 1491  
 Db 1515 CAGCTGAG 1574  
 Qy 1492 ATCAG 1551  
 Db 1575 GTCCAAATCAG 1634  
 Qy 1552 CTAAAGATTTGAG 1611  
 Db 1635 CTGAGAGATTTAG 1694  
 Qy 1612 GGAAGAGATTTAG 1671  
 Db 1695 ATCTATCTGCTGCTGAG 1754  
 Qy 1672 AGTTGCAATAG 1731  
 Db 1755 AACACACTGATTCAG 1814  
 Qy 1732 CGGAG 1791  
 Db 1815 TGCCAAAGCTTTAAG 1874  
 Qy 1792 GAGATGATGTTTCAACAG 1851  
 Db 1875 GAATGATGATTTAATCAATCAG 1934  
 Qy 1852 CTCAG 1911  
 Db 1935 TTAGCCTTTGAACCACTTCAATTAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1994  
 Qy 1912 AGCTGAG 1934  
 Db 1995 AGATTAG 2017

RESULT 2  
 US-09-879-957-193  
 ; Sequence 193, Application US/09879957  
 ; Patent No. US20020034755A1  
 GENERAL INFORMATION:  
 APPLICANT: SPARKS, Andrew B.  
 HOFFMAN, No. US20020034755A1h  
 KAY, Brian K.  
 FOWLES, Dana M.  
 MCCONNELL, Stephen J.  
 TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL  
 DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
 USING SAME  
 NUMBER OF SEQUENCES: 227  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Pennie & Edwards LLP  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10036-2711  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/09/879,957
FILING DATE: 13-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/630,915
FILING DATE: 03-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-5090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 193:
SEQUENCE CHARACTERISTICS:
LENGTH: 2873 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 193:
US-09-879-957-193

Query Match          9.1%; Score 332.4; DB 10; Length 2873;
Best Local Similarity 61.0%; Pred. No. 4.6e-75;
Matches 631; Conservative 0; Mismatches 361; Indels 42; Gaps 4;

QY 2609 CTTCTCTGACCGTACTGAGTGTGGCCAGTTACGGCAGAGATCAGCCTTTACCCCGCA 2668
DB 541 CTTTTCACAACTTACTCTAATATCATCATGCGCAAAATAATCAGCCTTCACCGCACTG 600
QY 2669 CAGCCACGTGCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 2728
DB 601 TGTCCCTCG--ATCTGTATCACCATTATTTATGACACGGACAAAGTGATGAAAACTTAA 657
QY 2729 AAGCGCAAGCCCTGTATCCCTGAGAGCCAAAAAGACAACAACCTTAAATTTTACAAA 2788
DB 658 AAGCAGAGGCCCTTGTCTGAGCTGCAAAAGAAATACCACTTGAACTTCAAAAC 717
QY 2789 GTGACGTCATCACCCTTCTGGAACAGACACATGTGTGTGTTGAGAACTTCAAGGTC 2848
DB 718 ATGACATTATTAAGTCTTGAGACAGCAAAATGTGTGTGAGGTGACATGAGAG 777
QY 2849 AGAAGGTTGTTCCCAAGCTTAACTGTAAGTCAATTCAGGCGCCGTAAAGAAATCCA 2908
DB 778 GAAAGAGATGTTTCCAAATCTTATGTCAAGATCATTTCTTGAGAGTAAATAAACGGG 837
QY 2909 CAAGCATCGATCTGCGCCCTACTGAAAGTCTCTAGTCTTAAAGAGAGTGGCTTCCCGG 2968
DB 838 AA-----GAACCAAGAGCTTTGTATGACGTGTAATAAAGAAACCTAACCTCGG 885
QY 2969 CCGGCAAGCCAGCTTCCCGGAGAAAGTTTATGTCATGTACATACATACAGAGATTCTG 3028
DB 886 CACCTTATTAAGTT-----GGAGAAAGAAATATTTGACCTTTATCAATTTCAAGGTGG 939
QY 3029 AGCAGAGAGTTTAACTTACAGCAAGAGGATGATGATGTTGTTGTTACCAAGAAAGATGGT 3088
DB 940 AACCTGAGATTGACTTCAAGAAAGTGAAGAAATATTTGGTGAACCCAGAAAGATGGAG 999
QY 3089 ACTGTGAGAGGAAACGGTGGGCGACAAGTCCGAGCTTCCCTTCAATATATGATGAGC 3148
DB 1000 AGTGTGTGAGAGAGATTTGGAGATGGAAGTGAATTTTTCATCAAACTATGTCAAC 1059
QY 3149 TTAAGATTCAAGAGGCTGTGAATCTGTGGGAAAAACAGGAGCTTTAAGAAAAAACTTG 3208
DB 1060 CAAAGGATCAAGAGAGTTTGGGAGCTGTAGCAAGTCTGAGCATCAATTAATAAAACCTG 1119
QY 3209 AAATTTGCCAGCTTATGCTTCTTCAAGCTGTACTGTCTCCGAAACAATCAACCTGCGTC 3268
DB 1120 AGATTGTCTAGATTAATTCAGCATATGTCTTGTGTTCTGAAACAACCTTAAGCTTGGAC 1179
QY 3269 CTGGGAGAGTATTTCTGATCCGAAAAAGAACCCAGGTGATGGTGGAGAGAACTGC 3328

```

```

DB 1180 CAGACAGTTATATTAATTTAAAGAAAAATACAAATGGGTGGTGCAGAGAGATTAC 1239
QY 3329 AAGCTCAGGAAAAAGCGCAGATAGGTTGTTCCAGCAAAATATGTAACCTTTAA 1289
DB 1240 AGGCCAGAGAAAAAAGCGACAGAAAGATGTTTCTGCGACGATCTTAACCTTTGG 1299
QY 3389 GCCCGGAACAAGCAAAATCAACCCCACTGAGCTAACCCAAAGCCGAGTGCAGCCAG 1348
DB 1300 GTCCAAAGTAGAAAGAGCCACACCTGCTTTATC-----CTG 1338
QY 3449 TGTGCGAGTATGAGGATGTAGCATTTACACCGCCAGAAAGATGAGAACTAGCTTCA 1398
DB 1339 TATGTGAGTATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1398
QY 3509 GCAAGGCGCATATCATCAAGCTCTCAACAGAGAGAGCCGAGCTGTGAAAAAGAGAG 1368
DB 1399 CCAAGGCAACCTATTAATTTATGAAACAAAGATGATCTGATTTGTTGCGAGAGAGA 1458
QY 3569 TCAGTGGCAAGTTGGGCTCTTCCCATCAATTTATGTAAAGCTGACACAGATGAGAC 1428
DB 1459 TCAAGGGGTGATGCTGTCTCTTCTTCAAACTAGTAAAGATGACAGACATCAGATC 1518
QY 3629 CAGCCAGCAATGA 3642
DB 1519 CAAGTCAACAGTGA 1532

RESULT 3
US-09-764-868-125
; Sequence 125, Application US/09764868
; Patent No. US2002016871A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OR INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 125
; LENGTH: 4210
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-868-125

Query Match          9.1%; Score 331.4; DB 9; Length 4210;
Best Local Similarity 61.0%; Pred. No. 1e-74;
Matches 630; Conservative 0; Mismatches 361; Indels 42; Gaps 4;

QY 2609 CTTCTCTGACCGTACTGAGTGTGGCCAGTTACGGCAGAGATCAGCCTTTACCCAGCA 2668
DB 918 CTTTTCACAACTTACTGTAATATCATCATGCGCAAAATAATCAGCCTTCACTGAATCG 977
QY 2669 CAGCCACGTGCTCTCCCATCTCCCGTCTGGGCGAGGTTGAAAGGTGGAAGGCTAC 2728
DB 978 TGTCCCTCG---ATCTGTATCACCATTATTCATGACACGGGCAAGTGTGAAAACTTAA 1034
QY 2729 AAGCGCAAGCCCTGTATCCCTGAGAGCCAAAAAGACAACACTTAAATTTAACAATA 2788
DB 1035 AAGCAGAGGCCCTTGTCTGAGACTGCAAGAAAGATTAACACTTGAACTTCAAAAC 1094
QY 2789 GTGAGCTATCAGCTGTGTGGAACAGCAAGCATGTGTGTGTTGAGAGATTTAAAGTC 2848
DB 1095 ATGACATTATTAATGCTTTGAGAGCAAGAAATAATGTGTGTTGGGAGGTGATGAGAG 1154
QY 2849 AGAAGGTTGTTGCCCAAGCTTAACTGTAAGTCAATTTAGGCGCCGTAAAGAAATCCA 2908
DB 1155 GAAAGAGATGTTTCCCAATCTTATGTCAAGATCATTTCTCGGAGAGTGAAGTAAAGGG 1214
QY 2909 CAAGCATGATATGCGCCCTTACTGAAAGTCTGTAGTCTTAAAGAGATGCTTCCCGG 2968

```

Db 1215 AA-----GAACCAAGCTTTGTATGACGCTGTAATAAGAACTACCTCG 1262  
 QY 2969 CCGGCAAGCCGATCCCGGAGAAAGATTATGCGACATGACATAGAGATTCTG 3028  
 Db 1263 CAGCTTATCGATT-----GGAGAGAAATATATTGCACTTTATCCATATTCAAGGTGG 1316  
 QY 3029 AGCAAGAGATTTTAACTTTTCAGCAAGGGGATGTGATTGTGTTTCCAGAAAGATGTG 3088  
 Db 1317 AACCTGGAGATTGACCTTTTCACAGAGGTGAAGAAATATTGGTGACCCAGAAAGATGGAG 1376  
 QY 3089 ACTGTGACGCGAAACGGTGGGCGACAAGTCCGAGTCTTCCCTTCTAATCTATGTGAGGC 3148  
 Db 1377 AGTGTGACAGGAAGATTTGAGATGAAAGTGAATTTTTCATCAACTAATGTCAAAC 1436  
 QY 3149 TTAAGATTTCAGAGGCTCTGGAACCTGCGGAAACAGGGAGTTTGAAGAAAAAAGCTG 3208  
 Db 1437 CAAGAGATCAAGAGATTTTGGAGTCTGCAAGCTGCGAGATCAATTAATAAAAGCTG 1496  
 QY 3209 AATTTCCGAGTTATTTGCTTCTAGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3268  
 Db 1497 AGATTGCTCAGGTAACTTCAACATATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1556  
 QY 3269 CTGGGAGCTGATTTCTGATCCGAAAAAGAACCCAGGTGATGTTGGAGAGAGAACTGC 3328  
 Db 1557 CAGAGACATTAATTTATTTTAAAGAAAAATACAGTGGGTGTGGCAAGAGAGATTAC 1616  
 QY 3329 AAGCTCAGGAAAAAGCGCCAGATAGGGTGTGTTCCAGCAAAATTATGTCAAACCTTTAA 3388  
 Db 1617 AGGCAAGAGAAAAAGCGAGAGAAAGATGTTCTTCCGACATGATTAACTTTTGG 1676  
 QY 3389 GCGCCGAGCAAGCAAAATACCCCACTGAGTACCCAGAACCGGAGTGCAGCCAGCAG 3448  
 Db 1677 GTCCAGATGTAAGAAAGCCACACCTGCTTTTCATC-----CTG 1715  
 QY 3449 TGTGCGCAGTGTGCGGATGTACGATTACACCGCCAGAAAGATGACGAAGTACCTTCA 3508  
 Db 1716 TATGTGAGTATTTGCTATGTATGATGATGAGCAATTAATGAAGATGAGCTCAGTTCT 1775  
 QY 3509 GCAAGGCGCATATCATACGCTCTCAACAGAGAGACCGGAGTGTGAAAGAGAGAG 3568  
 Db 1776 CCAAGGAGCAACTATTAATGTATGAACAAAGATGATCTGATTGTGCAAGGAGAGA 1835  
 QY 3569 TCAGTGGGCAAGTGGGCTCTTCCATTCATTTATGTAAGCTGACACAGAGATGGACC 3628  
 Db 1836 TCAACGGGGTACGTGCTCTTTCTTCAACTACGTTAAGATGACGACAGACTCAGATC 1895  
 QY 3629 CCAGCCAGCAATG 3641  
 Db 1896 CAAGTCAACAGTG 1908

RESULT 4  
 US-09-879-957-39  
 Sequence 39, Application US/09879957  
 Patent No. US2002003475A1  
 GENERAL INFORMATION:

APPLICANT: SPARKS, Andrew B.  
 HOFEMAN, No. US2002003475A1h  
 KAY, Brian K.  
 FOWLER, Dana M.  
 MCCONNELL, Stephen J.

TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL  
 DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
 USING SAME

NUMBER OF SEQUENCES: 227  
 CORRESPONDENCE ADDRESSES:

ADDRESSEE: Pennie & Edmonds LLP  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/879,957  
 FILING DATE: 13-Jun-2001  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/630,915  
 FILING DATE: 03-APR-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Mistic, S. Leslie  
 REGISTRATION NUMBER: 18,872  
 REFERENCE/DOCKET NUMBER: 1101-174  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 790-9090  
 TELEFAX: (212) 869-8864/9741  
 INFORMATION FOR SEQ ID NO: 39:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 747 bases  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA  
 SEQUENCE DESCRIPTION: SEQ ID NO: 39:  
 US-09-879-957-39  
 Query Match 7.1%; Score 258; DB 10; Length 747;  
 Best Local Similarity 63.9%; Pred. No. 2,8e-56;  
 Matches 418; Conservative 0; Mismatches 215; Indels 21; Gaps 1;  
 QY 2989 GGAGAGAGATTATTTCCATGTACATACAGAGTTCTGAGCAAGAGATTAACTTT 3048  
 Db 115 GGAGAGAGATTATTTCCATTTATTCATATTCAGTGTGAACTGAGATTGACTTTTC 174  
 QY 3049 CAGCAAGGAGTGTGATTTGTTGTTTACCAAGAAAGATGTGCTGTGAGCGGAGACGGT 3108  
 Db 175 ACAGAGGTGAAGAAATTTGTGTGCCAGAAAGATGTGAGTGTGACAGGAATTT 234  
 QY 3109 GAGCAAGTCCGAGTCTTCCCTTCTACTATGTGAGGCTTTAAAGATTGAGAGGCTCT 3168  
 Db 225 GAGATTAAGAGAAATTTTCCATCAATATGTCAACCAAGATCAAGAGATTTT 294  
 QY 3169 GGAATGCTGGGAAAAACGAGATTTAGAAAAAACTGAAATTTGCCAGATTATGCT 3228  
 Db 295 GGAATGCTACCAAGTCTGAGCATCAATTAATAAAACCTGAGATTGCTCAGTTACTTCA 354  
 QY 3229 TCCTACGCTGCTACTGTGTCGCGAACAATCACTCACTGCTCCTGGCAGCTGATTCTGATC 3288  
 Db 355 GCATATGTTGCTTCTGCTGTGTAACAACTTAGCTTGCACAGGACGTTATTTATTT 414  
 QY 3289 CGGAAAAAGAACCCAGGTGATGTGTGGAGAGAGAACTGCAAGCTGCAAGGAAAAAGCC 3348  
 Db 415 CTAAAGAAAAATATCAAGGTGTGTGGCAGAGAGATTACGCGCAAGAGAAAAACGA 474  
 QY 3349 CAGATAGGAGTGTTCAGCAAAATTTATGTCAAACTTCAAGCCCGGAACAGCAATATC 3408  
 Db 475 CAGAAAGATGTTTCTGCGCATCATGTTAACTTTTGGGTCCAGACAGTGAAGAGCC 534  
 QY 3409 ACCCAACTGAGTACCCCAAGACCGCAGTGCAGCAGAGTGTGCGAGGTGATGGAGT 3468  
 Db 535 ACACCTGCTTTCATCTT-----GTATGTCAAGGTGATGCTATG 573  
 QY 3469 TACGATTACACCGCCCAAGATGAGAACTAGCTTTCAGCAAGCCGACGATCATGAC 3528  
 Db 574 TATGACTATGAGCAAAATTAAGATGAGTCACTTTTCCAGAGCAACTATTATAT 633  
 QY 3529 GTCTCAACAAGAGAGACCGGAGTGTGCAAGAGAGAGTCACTGTGGCAAGTTGGGCTC 3588  
 Db 634 GTTATGAACAAAGATGATCTGATTGGTGGCAAGAGAGATCAACGGGGTGAAGTGTGCTC 693

QY 3589 TTCCCATTCATTATGTAAAGCTGACACGACATGACCCCGACGACATGA 3642  
 DB 694 TTTCCTTCAAACTAGCTTAAGATGACGACGACTGAGATCCAAAGTCAACAGTGA 747

RESULT 5  
 US-09-864-761-17127  
 ; Sequence 17127, Application US/09864761  
 ; Patent No. US20020048763A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Penn, Sharon G.  
 ; APPLICANT: Rank, David R.  
 ; APPLICANT: Hanzel, David K.  
 ; APPLICANT: Chen, Wensheng  
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 ; FILE REFERENCE: Aecmca-X-1  
 ; CURRENT APPLICATION NUMBER: US/09/864,761  
 ; PRIOR FILING DATE: 2001-05-23  
 ; PRIOR APPLICATION NUMBER: US 60/180,312  
 ; PRIOR FILING DATE: 2000-02-04  
 ; PRIOR APPLICATION NUMBER: US 60/207,456  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: US 09/632,366  
 ; PRIOR FILING DATE: 2000-08-03  
 ; PRIOR APPLICATION NUMBER: GB 24263,6  
 ; PRIOR FILING DATE: 2000-10-04  
 ; PRIOR APPLICATION NUMBER: US 60/236,359  
 ; PRIOR FILING DATE: 2000-09-27  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: US 60/234,687  
 ; PRIOR FILING DATE: 2000-09-21  
 ; PRIOR APPLICATION NUMBER: US 09/608,408  
 ; PRIOR FILING DATE: 2000-06-30  
 ; PRIOR APPLICATION NUMBER: US 09/774,203  
 ; PRIOR FILING DATE: 2001-01-29  
 ; NUMBER OF SEQ ID NOS: 49117  
 ; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1  
 ; SEQ ID NO 17127  
 ; LENGTH: 270  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: MAP TO AP000193.1  
 ; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.6  
 ; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.6  
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.3  
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6  
 ; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2  
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8  
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2  
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.6  
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.7  
 ; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2

OTHER INFORMATION: EST HUMAN HIT: AA773263.1, EVALUE 1.00e-112  
 ; OTHER INFORMATION: SWISSPROT HIT: Q15811, EVALUE 2.00e-45  
 ; OTHER INFORMATION: NT HIT: g11526214, EVALUE 0.00e+00  
 US-09-864-761-17127

Query Match 5.5%; Score 198.8; DB 10; Length 270;  
 Best Local Similarity 87.2%; Pred. No. 2.5e-41;  
 Matches 218; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 2299 GTGATGAAGACGACTGGAGAGCCAGATGGCTTGAGAGAGCTGAAGGAAAGACG 2358  
 DB 1 GTGATGAAGACGAACTGAGAGAACCCGCTGCTTGAGAGAGATTAAGAGAAAGACA 60  
 QY 2359 GGATGTTCCCTGCAAACTATGACGAAAGATTCAGAAAATGAGTTCCACTCCAGCC 2418  
 DB 61 GGGGTGTTCCCTGCAAACTATGACGAGAAAATCCAGAAAATGAGTTCCCGCTCCAGTG 120  
 QY 2419 AAACGATGACCATTCGACATCTGCCCTGCCCCCAACTGGCTGCTGAGACCCCT 2478  
 DB 121 AAACGATGACCATTCGACATCTGCCCTGCCCCCAACTGGCTGCTGAGACCCCT 180  
 QY 2479 GCTCCTTGGCCAGTACCTCTTCTGAGCCTCCGACAAACCCCAACTGGAGACTTC 2538  
 DB 181 GCCCTTGGCCAGTACCTCTTCTGAGCCTCCGACAAACCCCTTAATACCTGGCCGACTTC 240  
 QY 2539 AGTTCACGT 2548  
 DB 241 AGTTCACGT 250

RESULT 6  
 US-09-864-761-26948  
 ; Sequence 26948, Application US/09864761  
 ; Patent No. US20020048763A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Penn, Sharon G.  
 ; APPLICANT: Rank, David R.  
 ; APPLICANT: Hanzel, David K.  
 ; APPLICANT: Chen, Wensheng  
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 ; FILE REFERENCE: Aecmca-X-1  
 ; CURRENT APPLICATION NUMBER: US/09/864,761  
 ; PRIOR FILING DATE: 2001-05-23  
 ; PRIOR APPLICATION NUMBER: US 60/180,312  
 ; PRIOR FILING DATE: 2000-02-04  
 ; PRIOR APPLICATION NUMBER: US 60/207,456  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: US 09/632,366  
 ; PRIOR FILING DATE: 2000-08-03  
 ; PRIOR APPLICATION NUMBER: GB 24263,6  
 ; PRIOR FILING DATE: 2000-10-04  
 ; PRIOR APPLICATION NUMBER: US 60/236,359  
 ; PRIOR FILING DATE: 2000-09-27  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661  
 ; PRIOR FILING DATE: 2001-01-30  
 ; PRIOR APPLICATION NUMBER: PCT/US01/00670

PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: US 60/234,687  
 PRIOR FILING DATE: 2000-09-21  
 PRIOR APPLICATION NUMBER: US 09/608,408  
 PRIOR FILING DATE: 2000-06-30  
 PRIOR APPLICATION NUMBER: US 09/774,203  
 PRIOR FILING DATE: 2001-01-29  
 NUMBER OF SEQ ID NOS: 49117  
 SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
 SEQ ID NO: 26948  
 LENGTH: 286  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 OTHER INFORMATION: MAP TO AP000311.1  
 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.87  
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1  
 OTHER INFORMATION: EXPRESSED IN HELI00, SIGNAL = 1.3  
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 32  
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.66  
 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.69  
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1  
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.92  
 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.98  
 OTHER INFORMATION: NT HIT: AP114488.1, EVALUATION = 0.00e+00  
 OTHER INFORMATION: SWISSPROT HIT: Q15811, EVALUATION = 2.00e-45  
 OTHER INFORMATION: EST\_HUMAN HIT: AA773263.1, EVALUATION = 1.00e-112  
 US-09-674-761-26948

Query Match 5.5%; Score 198.8; DB 10; Length 286;  
 Best Local Similarity 87.2%; Pred. No. 2,6e-41;  
 Matches 218; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 2299 GTGATGAAAGCCAGCTGAGAGCCAGATGCTTGAGAGAGCTGAAAGGAAAGACG 2358  
 DB 1 GTGATGAAAGCCAGCTGAGAGCCAGATGCTTGAGAGAGCTGAAAGGAAAGACG 60  
 QY 2359 GGATGCTTCCCTGCAAACTATGACAGAAAGATTCAGAAATGAGTTCCCTCCAGCC 2418  
 DB 61 GGGTGTTCCTGCAAACTATGACAGAAAGATTCAGAAATGAGTTCCCTCCAGTG 120  
 QY 2419 AAACAGTGAACCGATGACATCTGCCCCCTGCCCCCAACTGCTGCTGAGACCCCT 2478  
 DB 121 AAACAGTGAACCGATGACATCTGCCCCCTGCCCCCAACTGCTGCTGAGACCCCT 180  
 QY 2479 GCTCTTTGGCAGTACCTCTTTCAGAGCCCTCCCAACCCCCCAACATGCGGAGACTTC 2538  
 DB 181 GCCCTTTGGCAGTACCTCTTTCAGAGCCCTCCCAACCCCCCAACATGCGGAGACTTC 240  
 QY 2539 AGTCCACGT 2548  
 DB 241 AGTCCACGT 250

RESULT 7  
 US-09-674-761-30453  
 Sequence 30453, Application US/09864761  
 Patent No. US20020048763A1

GENERAL INFORMATION:  
 APPLICANT: Penn, Sharon G.  
 APPLICANT: Han, David R.  
 APPLICANT: Hanzel, David K.  
 APPLICANT: Chen, Wensheng  
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 FILE REFERENCE: Aecmica-X-1  
 CURRENT APPLICATION NUMBER: US/09/864,761  
 PRIOR FILING DATE: 2001-05-23  
 PRIOR APPLICATION NUMBER: US 60/180,312  
 PRIOR FILING DATE: 2000-02-04  
 PRIOR APPLICATION NUMBER: US 60/207,456  
 PRIOR FILING DATE: 2000-05-26  
 PRIOR APPLICATION NUMBER: US 09/632,366

PRIOR FILING DATE: 2000-08-03  
 PRIOR APPLICATION NUMBER: GB 24263.6  
 PRIOR FILING DATE: 2000-10-04  
 PRIOR APPLICATION NUMBER: US 60/236,359  
 PRIOR FILING DATE: 2000-09-27  
 PRIOR APPLICATION NUMBER: PCT/US01/00666  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00667  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00664  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00669  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00665  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00668  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00663  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00662  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00661  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: PCT/US01/00670  
 PRIOR FILING DATE: 2001-01-30  
 PRIOR APPLICATION NUMBER: US 60/234,687  
 PRIOR FILING DATE: 2000-09-21  
 PRIOR APPLICATION NUMBER: US 09/608,408  
 PRIOR FILING DATE: 2000-06-30  
 PRIOR APPLICATION NUMBER: US 09/774,203  
 PRIOR FILING DATE: 2001-01-29  
 NUMBER OF SEQ ID NOS: 49117  
 SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
 SEQ ID NO: 30453  
 LENGTH: 297  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURE:  
 OTHER INFORMATION: MAP TO AP000117.1  
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.7  
 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4  
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2  
 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.76  
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3  
 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5  
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95  
 OTHER INFORMATION: NT HIT: AP114487.1, EVALUATION = 0.00e+00  
 OTHER INFORMATION: EST\_HUMAN HIT: AA773263.1, EVALUATION = 1.00e-112  
 OTHER INFORMATION: SWISSPROT HIT: Q15811, EVALUATION = 2.00e-45  
 US-09-674-761-30453

Query Match 5.5%; Score 198.8; DB 10; Length 297;  
 Best Local Similarity 87.2%; Pred. No. 2,7e-41;  
 Matches 218; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 2299 GTGATGAAAGCCAGCTGAGAGCCAGATGCTTGAGAGAGCTGAAAGGAAAGACG 2358  
 DB 1 GTGATGAAAGCCAGCTGAGAGCCAGATGCTTGAGAGAGCTGAAAGGAAAGACG 60  
 QY 2359 GGATGCTTCCCTGCAAACTATGACAGAAAGATTCAGAAATGAGTTCCCTCCAGCC 2418  
 DB 61 GGGTGTTCCTGCAAACTATGACAGAAAGATTCAGAAATGAGTTCCCTCCAGTG 120  
 QY 2419 AAACAGTGAACCGATGACATCTGCCCCCTGCCCCCAACTGCTGCTGAGACCCCT 2478  
 DB 121 AAACAGTGAACCGATGACATCTGCCCCCTGCCCCCAACTGCTGCTGAGACCCCT 180  
 QY 2479 GCTCTTTGGCAGTACCTCTTTCAGAGCCCTCCCAACCCCCCAACATGCGGAGACTTC 2538  
 DB 181 GCCCTTTGGCAGTACCTCTTTCAGAGCCCTCCCAACCCCCCAACATGCGGAGACTTC 240  
 QY 2539 AGTCCACGT 2548

Db 241 AGCTCCACGT 250

```

RESULT 8
US-09-864-761-17146
/ Sequence 17146, Application US/09864761
/ Patent No. US20020048763A1
/ GENERAL INFORMATION:
/ APPLICANT: Penn, Sharon G.
/ APPLICANT: Rank, David R.
/ APPLICANT: Hanzel, David K.
/ APPLICANT: Chen, Wensheng
/ TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
/ FILE REFERENCE: Aecm1ca-X-1
/ CURRENT APPLICATION NUMBER: US/09/864,761
/ PRIOR FILING DATE: 2001-05-23
/ PRIOR APPLICATION NUMBER: US 60/180,312
/ PRIOR FILING DATE: 2000-02-04
/ PRIOR APPLICATION NUMBER: US 60/207,456
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: US 09/632,366
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: GB 24263.6
/ PRIOR FILING DATE: 2000-10-04
/ PRIOR APPLICATION NUMBER: US 60/236,359
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: PCT/US01/00666
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00667
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00664
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00669
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00665
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00668
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00663
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00662
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00661
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00670
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: US 60/234,687
/ PRIOR FILING DATE: 2000-09-21
/ PRIOR APPLICATION NUMBER: US 09/608,408
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: US 09/774,203
/ NUMBER OF SEQ ID NOS: 49117
/ SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 17146
/ LENGTH: 301
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: MAP TO AP000311.1
/ OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
/ OTHER INFORMATION: EXPRESSED IN HB1100, SIGNAL = 3.2
/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3
/ OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
/ OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.3
/ OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.2
/ OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
/ OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.7
/ OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3
/ OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2
/ OTHER INFORMATION: SWISSPROT HIT: Q35601, EVALUATE 3.50e-01

```

US-09-864-761-17146

Query Match	5.5%;	Score 198.6;	DB 10;	Length 301;
Best Local Similarity	82.3%;	Pred. No. 3e-41;		
Matches 228;	Conservative 0;	Mismatches 49;	Indels 0;	Gaps 0;

Qy	1040	TGACATTGAAAGTAAGAAGCGGAGAACTTGTGACGAGGCTGTGGAGCTGGAGAAGC	1059
Db	1	TAACTGTTGAAGTAAGAAGCGGAGAACTTTGAACGTGCACTGGAACTGGAGAAGC	60
Qy	1100	GCCGCCAAGCGCTCTTTGGAGCAGCAGCGCAAAAGAGAGAGCGTTGGCTCACTGGAGC	1158
Db	61	GAAGGCAAGCTCTCTCTGGAAACAGCAGCGCAGAGACAGGAGCGCTGGCCAGCTGGAGC	120
Qy	1160	GGCGCCGAGCAGAGAGAAAGAGCGGAGCTGGCAGAGAGAGAGGCCCAAGCGCGACGTGG	1219
Db	121	GGGCGAGCAGAGAGAGAAAGAGACGTGAGCGCCAGAGAGCAAGAGCGCCAAAAGACAACTGG	180
Qy	1220	AGCTGAGAAAGCAGCTGGAGAAAGCAGCGGAGCTTGAGCGCGCAGCGAGAGAGAGAGA	1279
Db	181	AACGTGAGAAAGCAACTGGAAAAAGCAGCGGAGACTTGAAACGCGCAGAGAGAGAGAGAGA	240
Qy	1280	GGAAGGAGATCGAGAGGCGCGAGCGCCGCAAAAACGGGA	1316
Db	241	GGAAGAAATTGAGAGGCGAGAGGTTAAGCAGGCGAGGA	277

```

RESULT 9
US-09-864-761-17644
: Sequence 17644, Application US/09864761
: Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US 09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30

```







```
LENGTH: 487
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000049.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.3
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.3
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.9
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.5
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
US-09-864-761-864

Query Match
Best Local Similarity 84.6%; Score 180.8; DB 10; Length 487;
Pred. No. 1,4e-36;
Matches 203; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
```

```
QY 1039 GTGCAATTTGAAGTAAGAGCGGAGAACTTCAGCGAGGCACTGTGAGCTGAGAAAG 1098
DB 248 GTAAAGCTTTGAAGTAAGAGCGGAGAACTTTGAAGCTGCACTGAGAACTGAGAAA 307
QY 1099 CGCCGCAAGCGCTTTGAGAGCAGCGCAAGAGAGAGAGCGGTTGCTCAGCTGAGAG 1158
DB 308 CGAAGCAAGCTCTCTGGAACAGCAGCGCAAGAGAGAGAGCGGCTGCGCCAGCTGAGAG 367
QY 1159 CGCGCCGAGAGAGAGAGAGAGAGCGGAGAGCGGAGAGAGAGAGAGAGCGGAGAGAG 1218
DB 368 CGGCGGAGAGAGAGAGAGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 427
QY 1219 GAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1278
DB 428 GAACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 487
```

```
RESULT 13
US-09-864-761-311
Sequence 311, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecm1ca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIORITY APPLICATION NUMBER: US 60/180,312
PRIORITY FILING DATE: 2000-02-04
PRIORITY APPLICATION NUMBER: US 60/207,456
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: US 09/632,366
PRIORITY FILING DATE: 2000-08-03
PRIORITY APPLICATION NUMBER: GB 24263,6
PRIORITY FILING DATE: 2000-10-04
PRIORITY APPLICATION NUMBER: US 60/236,359
PRIORITY FILING DATE: 2000-09-27
PRIORITY APPLICATION NUMBER: PCT/US01/00666
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00667
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00664
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00669
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00665
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00668
```

```
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2001-01-29
PRIOR APPLICATION NUMBER: US 09/774,203
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 311
LENGTH: 475
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP000193.1
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.3
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
US-09-864-761-311
```

```
Query Match
Best Local Similarity 4.5%; Score 163; DB 10; Length 475;
Pred. No. 5.3e-32;
Matches 190; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
```

```
QY 2265 TGAGATCACATCCAGCCAGAGAGATATGATGCTGATGTAAGCCAGATGAGAGCC 2324
DB 241 TTAGATTTACTTAAGCGTACTTTTCCACAGGTGATGTAAGCCAACTGAGAAAC 300
QY 2325 AGATGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2384
DB 301 CGGCTGCTTGGAGAGAGAGATTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
QY 2385 AAAGATTCAGAAATGAGGTTCCCACTCCAGCCAAACCACTGACCGATTCACATCTGC 2444
DB 361 GAAATTCAGAAATGAGGTTCCCGCTCCAGTGAACCACTGACGTATTCACATCTGC 420
QY 2445 CCCTGCCCCCAAACAGCTGCTGCGTGAGAGACCCCTGCTTTCAGAGACCTCT 2499
DB 421 CCCTGCCCCCAAACAGCTGCTGCGTGAGAGACCCCGCCCTTTCAGAGACCTCT 475
```

```
RESULT 14
US-09-864-761-13884
Sequence 13884, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecm1ca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIORITY APPLICATION NUMBER: US 60/180,312
PRIORITY FILING DATE: 2000-02-04
PRIORITY APPLICATION NUMBER: US 60/207,456
```

PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 13864  
LENGTH: 475  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AP000117.1  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.7  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.76  
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95  
US-09-864-761-13884

Query Match 4.5%; Score 163; DB 10; Length 475;  
Best Local Similarity 80.9%; Pred. No. 5.3e-32;  
Matches 190; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 2265 TGAGATCACCATCCAGCCAGAGATATAGTCATGTGATGAAAGCCAGACTGAGAGCC 2324  
DB 241 TTAGTTCTACTTAAGCTGTGACTTTTTCACAGGTGATGAAAGCCAAACTGAGAAACC 300  
QY 2325 AGGATGGCTTGAGGAGAGGTGAAGGAGAGGATGTTCCCTCGCAACTATGCAACA 2384  
DB 301 CGGCTGGCTTGAGGAGAAATTAAGGAGAAAGACAGGAGGTGTTCCCTCGCAACTATGCAACA 360  
QY 2385 AAGATTCGAAATAGAGTTCCTCCACTCCAGCCAAACAGTACCGATCTGACATCTGC 2444  
DB 361 GAAATTCGAAATAGAGTTCCTCCACTCCAGCCAAACAGTACCGATCTGACATCTGC 420  
QY 2445 CCCTGCCCCCAAACTGGCTCTGCTGAGACCCCTGCTCTTGGCAGTACCTCT 2499  
DB 421 CCCTGCCCCCAAACTGGCTCTGCTGAGACCCCTGCTCTTGGCAGTACCTCT 475

RESULT 15

US-09-864-761-17125  
Sequence 17125, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aomicca-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
PRIOR FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 17125  
LENGTH: 180  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AP000192.1  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.3  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5  
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.2  
OTHER INFORMATION: NT HIT: g111526214, EVALUATE 1.00e-84  
OTHER INFORMATION: EST HUMAN HIT: BE871545.1, EVALUATE 2.00e-84  
OTHER INFORMATION: SWISSPROT HIT: Q50360, EVALUATE 2.20e-02  
US-09-864-761-17125

Query Match

4.0%; Score 146.4; DB 10; Length 180;

Best Local Similarity 88.3%; Pred. No. 5.7e-28;  
Matches 159; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

```

QY 347 GTATTAGGAGGATTGGCTAGCATGCCACCACTACAGCTGTGCTCCTGTCATGAGGCT 406
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1 GATGGAGGATTCGCCAGCATGCCACCGCTTACAGCTGTGCTCCAGTCCCATGGGAT 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 407 CCATTCCAGTTGTGGATGATGTCCACCCCTTAGTATCTTGTGTCCCTCCAGCAGATGC 466
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 CCATTCCAGTTGTGGATGATGTCCACCCCTTAGTATCTTGTGTCCCTCCAGCAGATGC 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 467 CTCCTGCTAGAGGGGCTCCCTCCGTCATAGACCTCTGCTGCTTGGGATCTCTG 526
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 CCCCCCTGGCTAAGGGGGCTCCCCCTGTATACCAACCTCTGCTGCAATTGCTCATCTG 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Search completed: March 14, 2003, 11:37:18  
Job time : 695.293 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2003, 01:40:35 ; Search time 4383.26 Seconds  
(without alignments)  
13456.658 Million cell updates/sec

Title: US-09-674-237A-2

Perfect score: 3642  
Sequence: 1 atggctcagttccacacac.....tggaccacgacgacatga 3642

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST: \*  
1: em\_estba: \*  
2: em\_esthum: \*  
3: em\_estin: \*  
4: em\_estnu: \*  
5: em\_estov: \*  
6: em\_estopl: \*  
7: em\_estro: \*  
8: em\_hc: \*  
9: gb\_est1: \*  
10: gb\_est2: \*  
11: gb\_hc: \*  
12: gb\_est3: \*  
13: gb\_est4: \*  
14: gb\_est5: \*  
15: em\_estfun: \*  
16: em\_estcom: \*  
17: gb\_gse: \*  
18: em\_gse\_hum: \*  
19: em\_gse\_inv: \*  
20: em\_gse\_pin: \*  
21: em\_gse\_vrt: \*  
22: em\_gse\_fun: \*  
23: em\_gse\_mam: \*  
24: em\_gse\_mus: \*  
25: em\_gse\_other: \*  
26: em\_gse\_pro: \*  
27: em\_gse\_rod: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1496.6	41.1	2079	11	BC013578 Homo sapi
2	1495	41.0	2126	11	BC020269 Homo sapi
3	795.6	21.8	911	14	BO942708 AGENCOURT
4	722.8	19.8	732	14	BM950476 UI-M-EHOP
5	721.2	19.8	738	14	BM944544 UI-M-EHOP
6	658.2	18.1	979	14	BO719508 AGENCOURT

Result No.	Score	Query Match	Length	ID	Description
7	656.6	18.0	690	14	BO179493 UI-M-EWO-
8	633.6	17.4	650	10	BB656585 BB656585
9	606.8	16.7	610	14	BQ443095 UI-M-EVO-
10	602	16.5	603	9	A1594919
11	593.2	16.3	606	10	AM910367
12	563.4	15.5	773	14	BM681943
13	558.8	15.3	639	14	BQ201258
14	555	15.2	1062	14	BQ898718
15	537.8	14.8	724	10	BE373101
16	533.8	14.7	537	9	AA063751
17	530.4	14.6	966	12	BG297300
18	529	14.5	896	14	BQ941336
19	527	14.5	734	9	A1671143
20	509	14.0	521	9	A1303871
21	504.6	13.9	875	12	BF122946
22	489.6	13.4	866	10	BE587712
23	484.6	13.3	527	10	BB757493
24	480.8	13.2	600	9	AL134506
25	479.8	13.2	483	10	BE200514
26	473.2	13.0	962	12	BG249478
27	468.4	12.9	481	9	AA212646
28	468.2	12.9	641	14	BM728997
29	463.2	12.7	556	9	A1580905
30	463	12.7	1009	12	BE871545
31	461.6	12.7	676	10	BB650196
32	448.4	12.3	510	12	BE952536
33	444	12.2	1015	14	BQ940346
34	442.2	12.1	447	10	BE655487
35	441.4	12.1	574	12	BF816839
36	441	12.1	501	9	A1852070
37	433.2	11.9	1767	11	BC020921
38	432.6	11.9	527	9	AL120063
39	432.4	11.9	443	9	AA492783
40	422.8	11.6	546	10	AM668553
41	420.6	11.5	628	14	BQ393010
42	414	11.4	475	10	BE199657
43	410	11.3	503	12	BF918745
44	404	11.1	979	14	BQ941411
45	403.2	11.1	568	10	AM594325

## ALIGNMENTS

RESULT 1  
LOCUS BC013578  
DEFINITION Homo sapiens, Similar to intersectin 1 (SH3 domain protein), clone IMAGE:3878242, mRNA.  
ACCESSION BC013578  
VERSION BC013578.1 GI:15488896  
KEYWORDS HTC.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 2079)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (04-SEP-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK  
COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [gcgaps-remail.nih.gov](mailto:gcgaps-remail.nih.gov)  
Tissue Procurement: DCTD/DTP/Gazdar  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: The I.M.A.G.E. Consortium (ILNL)  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www.sngc.stanford.edu>  
Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)

Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILN at: <http://image.llnl.gov>  
 Series: IRAX Plate: 14 Row: c Column: 5  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4808822  
 This clone has the following problem: incomplete processing.

## FEATURES

Location/Qualifiers  
 1..2079  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3878242"  
 /tissue\_type="Lung, large cell carcinoma"  
 /clone\_id="NIH MGC\_68"  
 /lab\_host="DH10B"  
 /note="Vector: PCMV-SPORT6"  
 BASE COUNT 703 a 451 c 492 g 433 t  
 ORIGIN

Query Match 41.1%; Score 1496.6; DB 11; Length 2079;  
 Best Local Similarity 86.0%; Pred. No. 0;  
 Matches 1671; Conservative 0; Mismatches 269; Indels 3; Gaps 1;

QY 1 ATGCTCAGTTCCTCCACACCTTTGGTGGTACCTGTGATGTCGGCCATTACTGTGGAG 60  
 Db 132 ATGGCTCAGTTCCTCCACACCTTTGGTGGTACCTGTGATGTCGGCCATTACTGTGGAG 191  
 QY 61 GAAAGGGCAGCATGACACAGCTTCTAGCTGAAGCCGATAGCGGATTTACT 120  
 Db 192 GAAAGGGCAGCATGACACAGCTTCTAGCTGAAGCCGATAGCGGATTTACT 251  
 QY 121 GGTATCAGCGAGAACTTTTTCATCTGGTTTACTGAGCTGTCTTACACAA 180  
 Db 252 GGTATCAGCGAGAACTTTTTCATCTGGTTTACTGAGCTGTCTTACACAA 311  
 QY 181 ATATGGGCGCTAGCGGACATGAATACGATGAGATGATCAAGTGAATTTTCCATA 240  
 Db 312 ATATGGGCGCTAGCGGACATGAATACGATGAGATGATCAAGTGAATTTTCCATA 371  
 QY 241 GCCATGAAGCTTATCAACTGAAGCTACAAAGATATCAGCTCCCTCCACACTTCCCT 300  
 Db 372 GCTATGAAGCTTATCAACTGAAGCTACAAAGATATCAGCTCCCTCCACACTTCCCT 431  
 QY 301 GTCATGAAACGACACAGTGGCTATTTCCAGTCCACAGATTTGGTATAGAGGAT 360  
 Db 432 GTCATGAAACGACACAGTGGCTATTTCCAGTCCACAGATTTGGTATAGAGGAT 491  
 QY 361 GCTAGCATGACACACTCACAGCTGTGCTCTGTGCAATGGGCTCCATTCAGTTGT 420  
 Db 492 GCGAGCATGACACCGCTTACAGCTGTGCTCTGTGCAATGGGCTCCATTCAGTTGT 551  
 QY 421 GGAATGCTCCACCTTATGATATCTTGTCTCCCTCCACAGAGAGTCTCCCTGGCTAAC 480  
 Db 552 GGAATGCTCCACCTTATGATATCTTGTCTCCACAGAGAGTCTCCCTGGCTAAC 611  
 QY 481 GGGGCTCCTCCGTCATACAGCTCTGTGCTGCTGTTGGGCTCCTGTAGCCACATGGCCA 540  
 Db 612 GGGGCTCCTCCGTCATACAGCTCTGTGCTGCTGTTGGGCTCCTGTAGCCACATGGCCA 671  
 QY 541 AAGAGTCTTCTTCAGCAGATCTGTGTCAGAGGTACAAATTAACTAAAGTTACAGAG 600  
 Db 672 AAGAGTCTTCTTCAGCAGATCTGTGTCAGAGGTACAAATTAACTAAAGTTACAGAG 731  
 QY 601 GCAAAATCATTCGATGTCGCGAGGCGCCCTCCACAGACAGAAATGGGTGGCTCAGTCA 660  
 Db 732 GCAAAATCATTCGATGTCGCGAGGCGCCCTCCACAGACAGAAATGGGTGGCTCAGTCA 791  
 QY 661 TCAAGGTGAAATACAGCAGTATATCAACAGCCACGACAAATATAGATGAGCACTTA 720  
 Db 792 TCAAGGTGAAATACAGCAGTATATCAATAGTATGACAAATATAGATGAGCACTTA 851

QY 721 ACAGGTCCCAAGCGAAGAACTATTCATGCAATCAAGTTTACCCAGGCTCAGCTGCT 780  
 Db 852 ACAGGTCCCAAGCGAAGAACTATTCATGCAATCAAGTTTACCCAGGCTCAGCTGCT 911  
 QY 781 TCAATATGGAATCTTTGACATGATGATCAAGTGAAGAACTCAGTGCAGAAATTTATC 840  
 Db 912 TCAATATGGAATCTTTGACATGATGATCAAGTGAAGAACTTACAGAGAGAAATTTATC 971  
 QY 841 CTAGCTATGACCTTAATGATGTTGCCATGTCGTGTGAGCCACTGCGCCGTCTGCT 900  
 Db 972 CTGCGAATGACCTCATTTGATGATGATGCTGCTGCGCAACCACTGCGCCGTCTGCT 1031  
 QY 901 CCAGAAATATATCCCTCTCTTCTTCCAGAAAGATTTGGCTCCGCAATGGGATGCTCGTATA 960  
 Db 1032 CCAGAAATATATCCCTCTCTTCTTCCAGAAAGATTTGGCTCCGCAATGGGATGCTCGTATA 1091  
 QY 961 AGCTTTCTTCTGATGATCAGAGGCTGCTGAGAGCCGTGTGAGAGATGACAGCAG 1020  
 Db 1092 AGCTTCAATCTGATGATCAGAGGCTTACAGAGGACCAAGTTTAAAGATGAACAAACAA 1151  
 QY 1021 C--CAGAGAAAGAACTGCTGTGACATTTGAAGTAAAGCCGGAGAACTTTCAGCGA 1077  
 Db 1152 CAATTAGAAAGAAATTTCTGTAAAGTTTGAAGTAAAGCCGGAGAACTTTCAGCGA 1211  
 QY 1078 GGCAGTGTGAGCTGTGAGAGAGCGCCAGCGCTTGTGAGCAGCAGCCGAAAGAGCAG 1137  
 Db 1212 GGCAGTGTGAGCTGTGAGAGAGCGCCAGCGCTTGTGAGCAGCAGCCGAAAGAGCAG 1271  
 QY 1138 GAGCGTTTGTGCTCAGCTGAGCGCGCCAGAGAGAGAGAAAGCCGGAGCGCCAGAG 1197  
 Db 1272 GAGCGCTTGTGCTCAGCTGAGCGCGCGAGAGAGAGAGAGAGAGCGCCAGAG 1331  
 QY 1199 CAGAGGCCAAGCCGCTGAGAGCTGAGAGAGCAGCTGAGAGAGAGCGGAGAGCTGGAG 1257  
 Db 1332 CAAAGCCGAAAGCACTGAATGAGAGAGCAGCTGAGAGAGAGCGGAGAGCTGAG 1391  
 QY 1258 CGGACGAGAGAGAGAGAGAGAGAGAGATCGAGAGCGCCGAGCGCCGAAACCGGAA 1317  
 Db 1392 CGGACGAGAGAGAGAGAGAGAGAGAGAGATCGAGAGCGCCGAGCGCCGAAACCGGAA 1451  
 QY 1318 CTGGAAGGCGAGCGCAACTTGAATGGAACCGAACCGAGACAGGAATCTCTGAATCAG 1377  
 Db 1452 CTGGAAGGCGAGCGCAACTTGAATGGAACCGAACCGAGACAGGAATCTCTGAATCAG 1511  
 QY 1378 AGGAACAGAGAGCAG 1437  
 Db 1512 AGGAACAGAGAGCAG 1571  
 QY 1438 GAGTTAGAGAGCTTGAATGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1497  
 Db 1572 GAGTTAGAGAGCTTGAATGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1631  
 QY 1498 TGTGACATGCGAAGCCAG 1557  
 Db 1632 TGTGACATGCGAAGCCAG 1691  
 QY 1558 ATTGCTGAATCACCACCTTACAGAGCAGTGTGAGAGAGAGAGAGAGAGAGAGAGAG 1617  
 Db 1692 ATTGCTGAATCACCACCTTACAGAGCAGTGTGAGAGAGAGAGAGAGAGAGAGAGAG 1751  
 QY 1618 CTATTTCCAGAGAAACAGATCTCAGTACCAAGTTAAACAAAGTCCAGCAGAAAGTTTG 1677  
 Db 1752 CTATTTCCAGAGAAACAGATCTCAGTACCAAGTTAAACAAAGTCCAGCAGAAAGTTTG 1811  
 QY 1678 CATAGAGACTGCTCTTACCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1737  
 Db 1812 CATAGAGACTGCTCTTACCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1871  
 QY 1738 CAGCTCCGGAGCAGCTGAT 1797  
 Db 1872 CAGCTCCGGAGCAGCTGAT 1931  
 QY 1798 GATGTTTTCACACACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1857

Db 1932 GATATTTTCAATATATGAGTGAAGAACTAAGAAATACACATATAGCAACACTCCAG 1991  
 Oy 1858 AAGCAGAGTCCCTGGAGCAGCGCACTGAAGCAGAAAGAGAGAGAGAGAGCTTG 1917  
 Db 1992 AAGCAAAAGTCCATGAGAGCTGACACACTGAAGAAAGAAAGAGAGAGATCATATA 2051  
 Oy 1918 GAGTTAGAGAGCAAAAGAGAGA 1940  
 Db 2052 GAATTAGAAAAACAAAAA 2074

## RESULT 2

LOCUS BC020269 2126 bp mRNA linear HTC 19-DEC-2001  
 DEFINITION Homo sapiens, clone IMAGE:4899011, mRNA.  
 ACCESSION BC020269  
 VERSION BC020269.1 GI:17939664  
 KEYWORDS HTC.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 2126)  
 AUTHORS Strausberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (19-DEC-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 CONTACT: MGC help desk  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: ATCC

CDNA Library Preparation: Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Genome Sequence Centre,  
 BC Cancer Agency, Vancouver, BC, Canada  
 info@bcgsc.bc.ca

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Buterfeld,  
 Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,  
 Letecia Heiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo  
 Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven  
 Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline  
 Schein, Duane Smalish, Michael Smith, Lorraine Spence, Jeff Stott,  
 Michael Thorne, Miranada Tsai, Natasia van den Bosch, Jill Vardy,  
 George Yang, Scott Zyderdyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>  
 Series: IRAL Plate: 40 Row: n Column: 1  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 4504796  
 This clone has the following problem: no cloning site /  
 microdeletion.

## FEATURES

Location/Qualifiers  
 1..2126  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4899011"  
 /tissue\_type="Pancreas, epithelioid carcinoma"  
 /clone\_lib="NIH MGC 42"  
 /lab\_host="DH10B-R"  
 /note="Vector: pOTB7"

BASE COUNT 712 a 462 c 517 g 435 t

## ORIGIN

Query Match 41.0%; Score 1495; DB 11; Length 2126;  
 Best Local Similarity 85.9%; Pred. No. 0;  
 Matches 1670; Conservative 0; Mismatches 270; Indels 3; Gaps 1;  
 Oy 1 ATGGCTAGTTTCCACACCTTTGGTGTAGCTGTGAGGCTATTAAGTGTGAG 60  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||

Db 182 ATGGCTAGTTTCCACACCTTTGGTGTAGCTGTGAGGCTATTAAGTGTGAG 241  
 Oy 61 GAAAGGGCCAAAGCATGACACAGCACTTCTTACCTGAGCGGATAGCGGATTTATTA 120  
 Db 242 GAAAGGGCCAAAGCATGATGACAGCTTCCATAGTTTAAAGCAATATGTGATCATTA 301  
 Oy 121 GGTGATCAAGGAGAGAACTTTTTCATTCGAGTTTACTTACGCTGTCTTACGACA 180  
 Db 302 GGTATCAAGCTAAGAACTTTTTCATTCGAGTTTACTTACGCTGTCTTACGACA 361  
 Oy 181 ATATGGGCGCTAGCGGAGCATGAATGAATGAATGAATGAATGAATGAATGAATGA 240  
 Db 362 ATATGGGCGCTAGCGGAGCATGAATGAATGAATGAATGAATGAATGAATGAATGA 421  
 Oy 241 GCATGAAGCTTTATCAACTGAAGCTTCAAGATATGAGTCTCCCTGACACTTCCCT 300  
 Db 422 GCTATGAAGCTTTATCAACTGAAGCTTCAAGATATGAGTCTCCCTGACACTTCCCT 481  
 Oy 301 GTATGAAGCAAGCAAGCTGAGTATTTCCAGTGCACAGCACTTTGTTATGAAGGAT 360  
 Db 482 GTATGAAGCAAGCAAGCTGAGTATTTCCAGTGCACAGCACTTTGTTATGAAGGAT 541  
 Oy 361 GCTATGAAGCAAGCAAGCTGAGTATTTCCAGTGCACAGCACTTTGTTATGAAGGAT 420  
 Db 542 GCATGAAGCAAGCAAGCTGAGTATTTCCAGTGCACAGCACTTTGTTATGAAGGAT 601  
 Oy 421 GGAATGTCCTCCAGCTTATGATATCTTCTGCTCCCTGACAGAGTCTCCCTGCTAAC 480  
 Db 602 GGAATGTCCTCCAGCTTATGATATCTTCTGCTCCCTGACAGAGTCTCCCTGCTAAC 661  
 Oy 481 GGGGCTCTCCCTGATACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540  
 Db 662 GGGGCTCTCCCTGATACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 721  
 Oy 541 AAGGCTCTCTCCCTGATACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600  
 Db 722 AAGGCTCTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 781  
 Oy 601 GCACATCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660  
 Db 782 GCACATCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 841  
 Oy 661 TCAAGCTGAAATATGAGCAAGTATTTCAACAGCCACGACAAATATGATGATGAT 720  
 Db 842 TCAAGCTGAAATATGAGCAAGTATTTCAATGATCATGATGATGATGATGATGAT 901  
 Oy 721 ACAAGTCTCCCGAGGAGCAATCTCATGCAATCAAGTTTACCCAGGCTCACTGCT 780  
 Db 902 ACAAGTCTCCCGAGGAGCAATCTCATGCAATCAAGTTTACCCAGGCTCACTGCT 961  
 Oy 781 TCAATATGAATCTTTCTGACATTTGATGATGATGATGATGATGATGATGATGAT 840  
 Db 962 TCAATATGAATCTTTCTGACATTTGATGATGATGATGATGATGATGATGATGAT 1021  
 Oy 841 TTAGCTATGACCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900  
 Db 1022 CTGGCAATGACCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1081  
 Oy 901 CCAGAAATATCT 960  
 Db 1082 CCAGAAATATCT 1141  
 Oy 961 AGCTTTCT 1020  
 Db 1142 AGCTTTCT 1201  
 Oy 1021 C---CAGAGAGAACTGCTCTGATCATTTGAAGATGAAGAGCGAGAACTTGAAGCA 1077  
 Db 1202 CAATTAGAAAGAAATTAACCTGTAACGTTTGAAGTAAGAGCGGAGAACTTGAAGCGT 1261  
 Oy 1078 GCGAGTGTGAGCTGTGAAGAGCGCGCCCAAGCGCTTTGGAGCAGCGCGGAGAGAGAG 1137  
 Db 1262 GCGAGTGTGAGCTGTGAAGAGCGCGCCCAAGCGCTTTGGAGCAGCGCGGAGAGAGAG 1321



OY	1138	GAGCGGTTGGCTCAGCTGGAGCGCCGCAGACAGAGAAGAAAGACGGGAGCGCCAGAG	1197
Db	1322	GAGCGCTGCCCACTGCTGGAGCGCGCGCAGACAGAGAGAGAACGTGATGCGCCAGAG	1381
OY	1198	CAGGAGGCCAAGCGGCACTGGAGCTTGAAGAACCTTGAGAGACAGCGGAGCTGGAG	1257
Db	1382	CAAGAGCGGAAAAAGAACCTGGAACTGGAGGAAGCAACTGGAAAAGACGGGAGCTAGAA	1441
OY	1258	CGGACGAGAGAGAGAGAGAGAGAAAGAGATCTGAGAGGCGGAGCGCCAAACCGGGA	1317
Db	1442	CGGCGAGAGAGAGAGAGAGAGAGAAAGAAATTGAGAGGCGAGAGGCTTCATAACCGGAA	1501
OY	1318	CTTGAAAAGCGACGACTTGAATGGGAACGAAACCGGAGACAGGAATCTCTGAATCAG	1377
Db	1502	CTTGAAAAGCGAACGCAACTTGAATGGGAAACGGAATCGAAAGCAAGAACTTAATCA	1561
OY	1378	AGGAACAAGAGACAGAGAGGCGACCGGCTCTGGAAGGCGAAGAGGAAAGCTCTGGAGTTT	1437
Db	1562	AGAAACAAGAAACAGAGAGACATAGTTTGACTGGAAGCAAGAAAAAGACTTTGGAAATTT	1621
OY	1438	GAGTTAAGAGCTCTGAATGACAAAAAGCATCAGTAGAAGAAAACTTCAGATATCAGG	1497
Db	1622	GAAATTAGAAGCTCTAAATATATAAAAAAGCATCAACTTAGAAGGAAAACTTCAGATATCAG	1681
OY	1498	TGTGCACCTGGCAACCAGAGGGAAGAAATTGAGAGCGACAAAGCTGTGAGAGCTAGGA	1557
Db	1682	TGTGATTTACCAACCCAAAGGCAAGAAATTGAGAGCACAAACAATTTGAGAGTTGAGA	1741
OY	1558	ATTGCTGAATCACCACCTTACAGCAGCAGTTGCAGGAATCTCAGCAATGCTTGGAGAA	1617
Db	1742	ATTGCCGAATCACCACCTTACAGCAACAATTACAGGAATCTCAGCAATGCTTGGAGAA	1801
OY	1618	CTTATTTCCAGAGAAACAGATCTCAGTGCACAGTTAAACAAGTCAGAGAGACAGTTTG	1677
Db	1802	CTTATTTCCAGAAAAACAGATCTCAGTGCACATTTAAACAAGTTGAGAGAACAGTTTG	1861
OY	1678	CATPAGAGATCGCTCTTACCTCTCAAAAGAGCCTTGGAAACAAAGAGCTGCCCCGAG	1737
Db	1862	CACAGAGATTCACCTGTTACACTTAAAAAGAGCTTGAAGCAAAAGAACTGACTCGGAG	1921
OY	1738	CAGCTCCGGGAGCAGCTGGACGAGGTGGAGAGAGAGACAGGTCAAAGTCGAGAGATTT	1797
Db	1922	CACCTACGAGCAACACTGGATGAAGTGGAGAAAGAACTGATCAAACTACAGAGATT	1981
OY	1798	GATGTTTTCAACCAACAGCAGCTGAAGCACTGAGAGAGATCATAGCAAAACAGCAACTCAG	1857
Db	1982	GATATTTTCAATATATAGCTGTAAGGAACCTAAGAGAAATACATATAGCAAACTCCAG	2041
OY	1858	AAGCAGAGGTCCTTGGAGGCAAGCGCACTGAAGCAGAAAGAGCAGAGAGAGAAAGCTTG	1917
Db	2042	AAGCAAAATCTCATGAGAGGCTGAACGACACTGAAACGAAAGAAACGAAGAAATCATTA	2101
OY	1918	GAGTTAGAGAACCAAAAGSAGA 1940	
Db	2102	GAAATTAAGAAAAAAGAAAAA 2124	
RESULT 3	BQ942708	911 bp mRNA linear EST 21-AUG-2002	
LOCUS	BQ942708		
DEFINITION	AGENCOURT 8764676 NIH_MGC_129 Mus musculus cDNA clone IMAGE:6314690		
ACCESSION	BQ942708	5' mRNA sequence.	
VERSION	BQ942708.1	GI:22358186	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
TITLE	Nih-MGC http://mhc.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC)		

Query Match	Best Local Similarity	21.8% ; Score 795.6 ; DB 14 ; Length 911 ;
Matches 890 ;	Conservative 0 ; Mismatches 9 ; Indels 9 ; Gaps 8 ;	
Qy 678	GCAGTTATTCACGACGACGACAAACTATGATGAGACACTTAAACAGGTCCCGACGCAAG	737
Db 1	GCAGTTATTCACGACGACGACAAACTATGATGAGACACTTAAACAGGTCCCGACGCAAG	60
Qy 738	AACATTCTCATGCAATTAAGTTATCCCGACGCTCAGTGGCTTCAATATGGAATCTTTC	797
Db 61	AACATTCTCATGCAATTAAGTTATCCCGACGCTCAGTGGCTTCAATATGGAATCTTTC	120
Qy 798	TGACATTGATCAAGATGAGAAATCTCAGTCAGAGAAATTTATCTAGCTATGACCTAAT	857
Db 121	TGACATTGATCAAGATGAGAAATCTCAGTCAGAGAAATTTATCTAGCTATGACCTAAT	179
Qy 858	TGATTTGTCATGTCGTGTGTCACGACATGTCGCCGCTCTGCTCCAGATATCATTCCTCC	917
Db 180	TGATTTGTCATGTCGTGTGTCACGACATGTCGCCGCTCTGCTCCAGATATCATTCCTCC	239
Qy 918	TTTCCTTCGAGAGATGTCGCTCCGAGGAGGATGTCGCTCAATAGCTCTTCTTGTTGGA	977
Db 240	TTTCCTTCGAGAGATGTCGCTCCGAGGAGGATGTCGCTCAATAGCTCTTCTTGTTGGA	299
Qy 978	TCAGAGCTGCTCTGAGAGCCGTCGTGACAGATGAGCAGCAGCCAGAGAAATCTGCC	1037
Db 300	TCAGAGCTGCTCTGAGAGCCGTCGTGACAGATGAGCAGCAGCCAGAGAAATCTGCC	359
Qy 1038	TGTGACATTTGAATGAAGAGCGGGAGAACTTGGAGCGAGGACGTGTGAGCTGGAGAA	1097
Db 360	TGTGACATTTGAATGAAGAGCGGGAGAACTTGGAGCGAGGACGTGTGAGCTGGAGAA	419
Qy 1098	GCGCGCCCAAGCGCTCTTGAGAGCAGCAGCCGAAAGAGCAGGAGCGGTTGGCTCAGCTGA	1157
Db 420	GCGCGCCCAAGCGCTCTTGAGAGCAGCAGCCGAAAGAGCAGGAGCGGTTGGCTCAGCTGA	479
Qy 1158	GCGCGCCCAAGCAGAGAGAGAAAGCGGGAGCCGACAGAGAGAGAGGCGCAAGCGGCGAGCT	1217
Db 480	GCGCGCCCAAGCAGAGAGAGAAAGCGGGAGCCGACAGAGAGAGAGGCGCAAGCGGCGAGCT	539
Qy 1218	GAGAGCTGGAAGAGAGCTGAGAGAGCAGCGGAGACTGAGCGGAGCGAGAGAGAGAGAG	1277
Db 540	GAGAGCTGGAAGAGAGCTGAGAGAGCAGCGGAGACTGAGCGGAGCGAGAGAGAGAGAG	599
Qy 1278	GAGGAGAGGATCGAGAGCGCGAGGCCGCAAAACGGGACACTGGAAGGACGAGCAACT	1337

D <sub>b</sub>	600	GAGGAAGGAGATTCGAGAGCCCGAAGCCGCCAAACCGGGAATCTGGAAAGCAGGCACA	ACT	659
Q <sub>y</sub>	1338	TGAATGGGAAAC - GGAACCCGAGACACGAAATCTTCAATCAGAGAACAGAGCAGGAGG		1396
D <sub>b</sub>	660	TGAATTGGGAACGGGAAACCGAGACACGGAAATCTTGAATCAGAGGAAACAAGGCGTAGAGG		719
Q <sub>y</sub>	1397	GC - ACCGTGGCTCGAAGGGCAAGGAGAAGCTGTGAGTTTGAGTTAGAAAGCTGTGA		1454
D <sub>b</sub>	720	GCCACCCTGTGCTCGAAGACAGAGAGAACTCTGGAGTTTGAGTTCAAGCTCTGAA		779
Q <sub>y</sub>	1455	TGACAAAAAGCATCAGTACGTAAGGAAAACTTCAGATATCAAGT - GTGCATCGCAACCC		1513
D <sub>b</sub>	780	TGACAAAAGCATCAGCTAGAAAGAAAACTTCGGATATCAGTGGTCCATCGGCAACCC		839
Q <sub>y</sub>	1514	AGA - GGCAAGAAATTGAGAGCACG - AACAACTCTAGA - GAGCTAAGAAATTGC - TGAATC		1569
D <sub>b</sub>	840	AGAGGGCAAGAAATTGAGAGCACGAAACAAGCTCTAAAGAGCTAAAAAATGCTTTGAAATC		899
Q <sub>y</sub>	1570	ACCGACTT	1577	
D <sub>b</sub>	900	ACCCAATT	907	

RESULT 4	
BM950476	
LOCUS	
DEFINITION	BM950476 732 bp mRNA linear EST 14-MAR-2002
	UI-M-EHOP-buu-o-05-0-UI.r1 NIH_BMAP_EHOP Mus musculus cDNA clone
	IMAGE:5687260 5', mRNA sequence.

ACCESSION	BM950476	
VERSION	BM950476.1	GI:19434066
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	

REFERENCE	Eukariota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus. 1 (bases 1 to 732)
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D.

Email: [CGAPbs-remail.nih.gov](mailto:CGAPbs-remail.nih.gov)  
Tissue Procurement: Dr. James Lin, University of Iowa  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: MCC clone distribution information can be  
found through the I.M.A.G.E. Consortium/BLND at:  
<http://image.ln1.gov>

```

FEATURES
  Seq primer: PYX-5.
  (BMAP)
  This clone was contributed by the Brain Molecular Anatomy project
  Location/Qualifiers
    1..732
    source

```

```

/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5687260"
/clone_1ib="NH_EMAP_EH09"
/tissue_type="whole brain"
/dev_stage="embryo 18.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/ncore="Organ: brain; Vector: pYX-Asc; Site 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldi, Lemon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I, and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail, is CAGCCACGAC. This library was created for the

```

BASE COUNT	ORIGIN
208 a	172 c 201 g 148 t 3 others

University of Iowa Mouse Brain Molecular Anatomy Project (BMAP): "Gene Discovery in the Developing Mouse Nervous System", supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

Query Match	19.8%	Score 722.8;	DB 14;	Length 732;
Best Local Similarity	99.3%	Pred. No. 3.3e-16;		
Matches 724; Conservative	0;	Mismatches 5;	Indels 0;	Gaps 0;

QY 2817 AGACATGTTGGTGTGGAGAAATTCAAGTCAGAAGGTGGTTCCCAAGCTTACTG 2876  
| | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 4 ACATTNGTGCGTGTGGAGAAATTCAAGTCAGAAGGTGGTTCCCAAGCTTACTG 63

DY 2877 GAACTCATTTGAGGGCCCGTAAGGAATCACAAGCATGCATCTGGCCCTACTGAAG 2936  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 64 GAAACTCATTTGAGGGCCCGTAAGGAATCACAAGCATGCATCTGGCCCTACTGAAG 123

[illegible]

QY 2997 GTTATTGGCATGTACACATACGAGAGTTCTTGAGCAGAGATTAACTTTACCTTTGACCAAGG 3056  
Dd 184 GTTATTGGCATGTACACATACGAGAGTTCTTGAGCAGAGATTAACTTTACCTTTGACCAAGG 243

Qy	3057	GGATGTGATTGTGGTTACCAAGAAAGATGGTGA	CGGGACGGTGGGCGACA	3116
Db	244	GGATGTGATTGTGGTTACCAAGAAAGATGGTGA	CGGGACGGTGGGCGACA	303

OY 3117 GTCCGGAGTCTTCCTTTAACTATGTGAGCCTTAAGAATTCAAGAGGCCTTGGAACCTGC 3176  
  
Db 304 GTCGGAGTCTTCCTTTAACTATGTGAGCCTTAAGAATTCAAGAGGCCTTGGAACCTGC 363

---

[illegible][illegible]

Oy	3297	GAA	CCAGATGGATGTG	TGGGAAGGAACTC	CACAGCTCGA	GGGAAAAAGGCC	CCAGATTAGG	3356
Dd	484	GAA	CCAGATGGATGTG	GGAAGGAACTCA	AGCTCGA	GGGAAAAAGGCC	CCAGATTAGG	543

Accession	Gene	Protein	Length
3442	LOC100286643	LOC100286643	3457
Oy	3357	GTGGTTTCACACAAATTATGCAAACTTCTAAGCCCGCGAAACAACAAATACCCCAAC	3416
Db	544	GTGGTTTCACACAAATTATGCAAACTTCTAAGCCCGCGAAACAACAAATACCCCAAC	603

Accession	Gene	Accession	Gene
3417	TCAGCTACCCCAAGACCCGACGATGCGACCGACGATGTCGCCAGGAGTACGATGATTA	3476	TCAGCTACCCCAAGACCCGACGATGCGACCGACGATGTCGCCAGGAGTACGATGATTA
604	TCAGCTACCCCAAGACCCGACGATGCGACCGACGATGTCGCCAGGAGTACGATGATTA	663	TCAGCTACCCCAAGACCCGACGATGCGACCGACGATGTCGCCAGGAGTACGATGATTA

Accession	Gene	Protein	Accession	Gene	Protein
0Y3477	CACCGCCAGAACGATGACGAACCTAGCCCTACGCAAAAGCCAGATCATCAAGCTTCCAA	3536	0Y3477	CACCGCCAGAACGATGACGAACCTAGCCCTACGCAAAAGCCAGATCATCAAGCTTCCAA	3536
Db664	CACCGCCAGAACGATGACGAACCTAGCCCTACGCAAAAGCCAGATCATCAAGCTTCCAA	723	Db664	CACCGCCAGAACGATGACGAACCTAGCCCTACGCAAAAGCCAGATCATCAAGCTTCCAA	723

v7		557 CAAGGAGA 565
Db		724 CAAGGAGA 732

RESULT 5					
BM944544					
LOCUS					
	BM944544	738 bp	mRNA	linear	EST 14-MAR-2002
DEFINITION	U1-M-EH00-bvr-i-08-0-0-U1.r1 NIH BMAP EH00 Mus musculus cDNA clone				

IMAGE:5695975 5' mRNA sequence.  
 ACCESSION BM944544  
 VERSION BM944544.1 GI:19428129  
 KEYWORDS EST.  
 SOURCE house mouse.

ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 738)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-remail.nih.gov  
 Tissue Procurement: Dr. James Lin, University of Iowa  
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/BLNI at:  
 http://image.llnl.gov  
 This clone was contributed by the Brain Molecular Anatomy Project  
 (BMAP)  
 Seq primer: pyx-5.  
 Location/Qualifiers  
 1. 738  
 /organism="Mus musculus"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5695975"  
 /clone\_1b="NIH\_BMAP\_EH09"  
 /tissue\_type="whole brain"  
 /dev\_stage="embryo 18.5 dpc"  
 /lab\_host="DH10B (T1 phase resistant)"  
 /note="Organ: brain; Vector: pyx-Asc; Site 1: EcoR I;  
 Site 2: Not I; The library was constructed according to  
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured mRNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with an  
 oligo-dT primer containing a Not I site. Double stranded  
 cDNA was size selected according to mRNA size fraction,  
 ligated with EcoR I adaptor, digested with Not I, and then  
 cloned directionally into the pyx-Asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 is CAGCCACGAC. This library was created for the  
 University of Iowa Mouse Brain Molecular Anatomy Project  
 (BMAP). 'Gene Discovery in the Developing Mouse Nervous  
 System', supported by National Institute of Mental Health  
 (NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT 198 a 213 c 184 g 143 t  
 ORIGIN  
 Query Match 19.8%; Score 721.2; DB 14; Length 738;  
 Best Local Similarity 99.5%; Pred. No. 8.1e-166;  
 Matches 734; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 2310 CCAGACTGAGAGAGCCAGATGGCTTGGAGGAGAGCTAAAGGAGAGCGAGATGTTCCC 2369  
 1 CCAGACTGAGAGAGCCAGATGGCTTGGAGGAGAGCTAAAGGAGAGCGAGATGTTCCC 60  
 Db 1 CCAGACTGAGAGAGCCAGATGGCTTGGAGGAGAGCTAAAGGAGAGCGAGATGTTCCC 60  
 QY 2370 TGCACCTATGACAGAAAGATTCAGAAATAGAGTTCCACTCCAGCCAAACAGTAC 2429  
 61 TGCACCTATGACAGAAAGATTCAGAAATAGAGTTCCACTCCAGCCAAACAGTAC 120  
 Db 1 TGCACCTATGACAGAAAGATTCAGAAATAGAGTTCCACTCCAGCCAAACAGTAC 120  
 QY 2430 CGATCTGACATGCGCCCTGCGCCCAAGCTGCTCGGAGAGCCCTCTCTTTGCC 2489  
 121 CGATCTGACATGCGCCCTGCGCCCAAGCTGCTCGGAGAGCCCTCTCTTTGCC 180  
 Db 1 CGATCTGACATGCGCCCTGCGCCCAAGCTGCTCGGAGAGCCCTCTCTTTGCC 180  
 QY 2490 AGTACCTCTTGGAGCCCTTCCACCAACCCCAACACTGGGAGAGCTTCAAGTTCACG 2549  
 181 AGTACCTCTTGGAGCCCTTCCACCAACCCCAACACTGGGAGAGCTTCAAGTTCACG 240  
 Db 1 AGTACCTCTTGGAGCCCTTCCACCAACCCCAACACTGGGAGAGCTTCAAGTTCACG 240  
 QY 2550 GCCAGAGAGCTCAACAGAGAGCCAGAAAGGAGACAACTGGGATTCGTGGCGGCTACGCC 2609  
 241 GCCAGAGAGCTCAACAGAGAGCCAGAAAGGAGACAACTGGGATTCGTGGCGGCTACGCC 300  
 QY 2610 TTCTCTGACCGTACCTGATGTGCGCAGTTACGGCAGAGATTCAGCTTACCCAGGCAC 2669  
 1 TTCTCTGACCGTACCTGATGTGCGCAGTTACGGCAGAGATTCAGCTTACCCAGGCAC 2669

Db 301 TTCTCTGACCGTACCTGATGTGCGCAGTTACGGCAGAGATTCAGCTTACCCAGGCAC 360  
 QY 2670 AGCCACTGGCTCTCCCATATCCCGTCTGGGCGAGAGTGAAGAGGAGGCTTACA 2729  
 361 AGCCACTGGCTCTCCCATATCCCGTCTGGGCGAGAGTGAAGAGGAGGCTTACA 420  
 Db 1 AGCCACTGGCTCTCCCATATCCCGTCTGGGCGAGAGTGAAGAGGAGGCTTACA 420  
 QY 2730 AGCCAGAGCTCTGATCCCTGGAGAGCCAAAGAGACCACTTAATTTTAAACAAG 2789  
 421 AGCCAGAGCTCTGATCCCTGGAGAGCCAAAGAGACCACTTAATTTTAAACAAG 480  
 Db 1 AGCCAGAGCTCTGATCCCTGGAGAGCCAAAGAGACCACTTAATTTTAAACAAG 480  
 QY 2790 TGAGCTATCACCCTGTTGGAAGAGACATGTGTGTTGGAGAGTTCAAGTCA 2849  
 481 TGAGCTATCACCCTGTTGGAAGAGACATGTGTGTTGGAGAGTTCAAGTCA 540  
 Db 1 TGAGCTATCACCCTGTTGGAAGAGACATGTGTGTTGGAGAGTTCAAGTCA 540  
 QY 2850 GAAGGTTGTTCCCAAGTCTTACGTGAATCTATTTCAGGCGCCGTAAAGAAATCCAC 2909  
 541 GAAGGTTGTTCCCAAGTCTTACGTGAATCTATTTCAGGCGCCGTAAAGAAATCCAC 600  
 Db 1 GAAGGTTGTTCCCAAGTCTTACGTGAATCTATTTCAGGCGCCGTAAAGAAATCCAC 600  
 QY 2910 AAGCATGATCTGGCCCTACTGAAAGTCCGTAGCTAAAGAGAGTGGCTCCCGGC 2969  
 601 AAGCATGATCTGGCCCTACTGAAAGTCCGTAGCTAAAGAGAGTGGCTCCCGGC 660  
 Db 1 AAGCATGATCTGGCCCTACTGAAAGTCCGTAGCTAAAGAGAGTGGCTCCCGGC 660  
 QY 2970 CGCCAGAGCAGCCATTCGCGAGAGAGATTATTCAGTATACATACAGAGTTCTGA 3029  
 661 CGCCAGAGCAGCCATTCGCGAGAGAGATTATTCAGTATACATACAGAGTTCTGA 719  
 Db 1 CGCCAGAGCAGCCATTCGCGAGAGAGATTATTCAGTATACATACAGAGTTCTGA 719  
 QY 3030 GCAGAGATTTAACTTT 3047  
 720 GCAGAGATTTAACTTT 737  
 Db 1 GCAGAGATTTAACTTT 737

RESULT 6  
 LOCUS B0719508 979 bp mRNA linear EST 16-JUL-2002  
 DEFINITION AGNCOURT 8219793 Lupski, sympathetic\_trunk Homo sapiens cDNA clone  
 IMAGE:618111 5', mRNA sequence.  
 ACCESSION B0719508  
 VERSION B0719508.1 GI:21858405  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 979)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-remail.nih.gov  
 Tissue Procurement: Dr. James R. Lupski  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/BLNI at:  
 http://image.llnl.gov  
 Plate: LLM13583 row: C column: 24  
 High quality sequence stop: 643.  
 Location/Qualifiers  
 1. 979  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:618111"  
 /clone\_1b="Lupski\_sympathetic\_trunk"  
 /sex="male"  
 /tissue\_type="sympathetic trunk"  
 /dev\_stage="adult, 16 yr"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6 (Life Technologies); Site 1:  
 NotI; Site 2: SalI; cDNA made by oligo-dT priming.  
 Directionally cloned using the following adaptors:  
 5'-TCGACCCAGCGTCGCG-3' and  
 5'-GACTAGTTTAAATCGCAGAGCGCCGCTT(15)-3'. Size selected >

1 kb for average insert length 1.9 kb. This is a primary library, non-amplified. Library constructed by Life Technologies and donated by J. Lipski, M.D./Ph.D. (Baylor College of Medicine); available through Life Technologies."

BASE COUNT 294 a 247 c 218 g 217 t 3 others  
ORIGIN

Query Match 18.1%; Score 658.2; DB 14; Length 979;  
Best Local Similarity 84.2%; Pred. No. 2.6e-150;  
Matches 776; Conservative 0; Mismatches 141; Indels 5; Gaps 3;

291 ACTTCCCTGTCATGAAACAGACACAGTGGCTATTTCAGTGCACGACATTTGGTAT 350  
1 ACTTCCCTGTCATGAAACAGACACAGTGGCTATTTCAGTGCACGACATTTGGTAT 60  
351 AGAGAGGATTTGTAAGATGCCACATCTACAGTGTGCTCTGCGCAATGGGCTCAT 410  
61 GGGAGGATTCGCGAGATCCACGACCGCTTACGCTGTGCTCCAGTGCACATGGATCCAT 120  
411 TCCAGTTGTAAGATGTCACACCTTATATCTCTGCTCCAGTGCACGATGCTCC 470  
121 TCCAGTTGTAAGATGTCACACCTTATATCTCTGCTCCAGTGCACGATGCTCC 180  
471 CCTGGCTAAAGGGGCTCTCTCCGTCATACAGCCTGCTGCGTTTCCGATCTGCACG 530  
181 CCGGCTAAAGGGGCTCTCTCCGTCATACAGCCTGCTGCGTTTCCGATCTGCACG 240  
531 CACATGGCCAAAGATTTCTCTTCAGCAGATCTGCTCCAGGTCACATTTAAACCTTA 590  
241 CACATGGCCAAAGATTTCTCTTCAGCAGATCTGCTCCAGGTCACATTTAAACCTTA 300  
591 GTTACGAGAGGACATCATTCATGCTCCAGCGCCCTCCAGCAGCAGATGGGCTGT 650  
301 ATTACAAAGGACATCATTCATGCTCCAGCGCCCTCCAGCAGCAGATGGGCTGT 360  
651 GCTCAGTATCAAGGCTGAAATACAGGAGTTTATCAAGCAGCAGCAGATTTATGAG 710  
361 TCTCAGTATCAAGGCTGAAATACAGGAGTTTATCAAGCAGCAGATTTATGAG 420  
711 TGGACACTTACAGGTTCCCGAGCAAACTATTCTCATGCAATCAAGTTTACCCAGGC 770  
421 TGGACACTTACAGGTTCCCGAGCAAACTATTCTCATGCAATCAAGTTTACCCAGGC 480  
771 TCGAGTGGCTCAATATGGAATCTTCTGACATTCATCAAGATGAAAACTCACTGCAGA 830  
481 TCGAGTGGCTCAATATGGAATCTTCTGACATTCATCAAGATGAAAACTCACTGCAGA 540  
831 AGAATTATCTAGTACTGACCTTAATTGATGTTCGATGTCTGATCAAGCACTGCGCGC 890  
541 GGAATTATCTAGTACTGACCTTAATTGATGTTCGATGTCTGATCAAGCACTGCGCGC 600  
891 GGTCTGCTCCAGAAATACATCCCTCTCTTCAAGAAAGTTGCTCCGCGAGTGGAT 950  
601 TGTCTGCTCCAGAAATACATCCCTCTCTTCAAGAAAGTTGCTCCGCGAGTGGAT 660  
951 GTGCGTATAGTCTTCTCTGATGACAGAGCTGCTGAGAGCCGCTGCACAGAGA 1010  
661 ATCTGTATAGTCTTCTCTGATGACAGAGCTGCTGAGAGCCGCTGCACAGAGA 720  
1011 TGAGCAGCAGC---CAGAGAAAGAACTGCTGTGACATTTGAAG-ATAAGAGCGGAGA 1066  
721 TGAGCAGCAGCATTGAAAGAAATTAAGTTTACCTGTGAAAGATTAAGAGCGGAGA 780  
1067 ACTTGAAGCAGGAGTGTGAGCTGGAAGAGCGCGCAAGGCTCTTGGAGCAGCAGC 1126  
781 ACTTGAAGCAGGAGTGTGAGCTGGAAGAGCGCGCAAGGCTCTTGGAGCAGCAGC 839  
1127 GCAAGAGCAGAGCGGTTGGCTCAGCTGAGGCGCGCGAGCAGAGAGAAAGAGCGGG 1186  
840 GCAAGAGCAGAGCGGTTGGCTCAGCTGAGGCGCGCGAGCAGAGAGAAAGAGAGCT 899  
1187 AGCGCCAGAGCAGAGGCGCA 1208

Db 900 GACCCCAAGAACAGAGCGCAAA 921

# RESULT 7

LOCUS

BO179493 690 bp mRNA linear EST 30-APR-2002

UI-M-EW0-bww-k-03-0-UI.r1 NIH\_BMAP\_EW0 Mus musculus cDNA clone

IMAGE:5704058 5', mRNA sequence.

BO179493

BO179493.1 GI:20354985

EST.

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 690)

NIH-MGC

http://mgc.nci.nih.gov/

Unpublished (1999)

contact: Robert Strausberg, Ph.D.

Email: cgaabs-r@mail.nih.gov

Tissue Procurement: Dr. James Lin, University of Iowa

CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa

CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa

DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LENL at:

http://image.lnl.gov

This clone was contributed by the Brain Molecular Anatomy Project

(BMAP)

Seq primer: pyx-5.

Location/Qualifiers

1..690

/organism="Mus musculus"

/strain="C57BL/6"

/db\_xref="taxon:10090"

/clone="IMAGE:5704058"

/clone\_lib="NIH BMAP EW0"

/tissue\_type="whole brain"

/dev\_stage="embryo 15.5 dpc"

/lab\_host="DH10B (T1 phage resistant)"

/note="Organ: brain; Vector: pyx-Asc; Site: 1; Ecov: I; Site 2: Not I; The library was constructed according to

Bonaldi, Lennon and Soares, Genome Research, 6:791-806,

1996. Denatured mRNA was size fractionated on a 1% agarose

gel. First strand cDNA synthesis was primed with an

oligo-dT primer containing a Not I site. Double stranded

cDNA was size selected according to mRNA size fraction,

ligated with EcoR I adaptor, digested with Not I, and then

cloned directionally into pyx-Asc vector. The library tag

sequence located between the Not I site and the polyA tail

is GTGCTGGA. This library was created for the

University of Iowa Mouse Brain Molecular Anatomy Project

(BMAP). "Gene Discovery in the Developing Mouse Nervous

System", supported by National Institutes of Mental Health

(NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT 185 a 197 c 176 g 129 t 3 others

Query Match 18.0%; Score 656.6; DB 14; Length 690;  
Best Local Similarity 98.7%; Pred. No. 5.6e-150;  
Matches 681; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

2233 GTGATGTGATGAAAGCCAGACTGAGAGCCAGAGTGGCTTGAAGAGAGCTGAAGGG 2352  
1 GTGATGTGATGAAAG-CAGACTGAGAGCCAGAGTGGCTTGAAGAGAGCTGAAGGG 59  
2353 AAGAGGGATGTTCCCTGCAAACTATGCAAGAAAGATTCCAGAAATGAGTTCCACT 2412  
60 AAGAGGGATGTTCCCTGCAAACTATGCAAGAAAGATTCCAGAAATGAGTTCCACT 119  
2413 CAGGCAAAACGATGACGATCTGACATCTGCCCTGCCCCCAAACTGGCTTGGCTGAG 2472

```

Db      120  CAGGCCAAACGAGGACGATCTGACATCGCCCTGCCCCCAACTGGCTCGCGTAG 179
      2473  ACCCGTCTCTCTTGGCCAGGACCTCTTCTAGAGCCCTCCCAACCCCAACTGCGCA 2532
      180  ACCCGTCTCTCTTGGCCAGGACCTCTTCTAGAGCCCTCCCAACCCCAACTGCGCA 239
      2533  GACTTCAGTTCCACGTGGCCCGACGCTCAACGAGAAAGCGAAACGAACTGGAT 2592
      240  GACTTCAGTTCCACGTGGCCCGACGCTCAACGAGAAAGCGAAACGAACTGGAT 299
      2593  ACGTGGGCGGCTGACGCTCTCTGACCGTCTAGTCTGGCCAGTTAGCGGAGATCA 2652
      300  ACATGGGCGGCTGACGCTCTCTGACCGTCTAGTCTGGCCAGTTAGCGGAGATCA 359
      2653  GCCTTTACCCCGACGACGACCTGCTCTCCCATCTCCGCTCTGCGGCGAGGTGAA 2712
      360  GCCTTTACCCCGACGACGACCTGCTCTCCCATCTCCGCTCTGCGGCGAGGTGAA 419
      2713  AAGGTGAAAGGCTCAAGAGCGCAAGCCCTGTATCCCTGGAGAGCC-AAAAAGACACCA 2771
      420  AAGGTGAAAGGCTCAAGAGCGCAAGCCCTGTATCCCTGGAGAGCGCAAAAAAGACACCA 479
      2772  CTTAAATTTTAAACAAGATGACGTCATCACCGTTCTGGAACAGACATGTGTGTT 2831
      480  CTTAAATTTTAAACAAGATGACGTCATCACCGTTCTGGAACAGACATGTGTGTT 539
      2832  TGGAGAAAGTTCAAGTCAAGAAAGGTTGTTCCCAAGTCTTACGTGAATCTTCAAG 2891
      540  TGGAGAAAGTTCAAGTCAAGAAAGGTTGTTCCCAAGTCTTACGTGAATCTTCAAG 599
      2892  GCGCGTAAAGAAATTCACAAGCATGATGATGCGCCCTACTGAAAGTCTCTGTCTTAA 2951
      600  GCGCGTAAAGAAATTCACAAGCATGATGATGCGCCCTACTGAAAGTCTCTGTCTTAA 659
      2952  GAGAGTGGCTTCCCGGCGCGCAAGCCAGC 2981
      660  GAGAGTGGCTTCCCGGCGCGCAAGCCAGC 689

RESULT 8
BB656585 650 bp mRNA linear EST 26-OCT-2001
LOCUS BB656585 RIKEN full-length enriched, 12 days embryo spinal ganglion
DEFINITION BB656585 Mus musculus cDNA clone D130043B15 5', mRNA sequence.
ACCESSION BB656585
VERSION BB656585.1 GI:16490413
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 650)
Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A.,
Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J., Kono, H., Kouda
, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, M., Ohno, M.,
Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki
, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,
Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T.,
Muramatsu, M., and Hayashizaki, Y.
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenho-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh
, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.

```

```

FEATURES
source
    Normalization and subtraction of cap-trapper-selected cDNAs to
    prepare full-length cDNA libraries for rapid discovery of new
    genes. Genome Res. 10 (10), 1617-1630 (2000)
    wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
    Marahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura
    , S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
    Hayashizaki, Y.
    RIKEN integrated sequence analysis (RISA) system--384-format
    sequencing pipeline with 384 multicapillary sequencer. Genome Res.
    10 (11), 1757-1771 (2000)
    Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara
    , Y. and Hayashizaki, Y.
    Computer-based methods for the mouse full-length cDNA
    encyclopedia: real-time sequence clustering for construction of a
    nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
    Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamataka, T., Aizawa
    , K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and
    Hayashizaki, Y.
    Computational Analysis of Full-Length Mouse cDNAs Compared with
    Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
    Please visit our web site (http://genome.gsc.riken.go.jp) for
    further details.
    e mouse tissues.
    Location/Qualifiers
    1..650
    /organism="Mus musculus"
    /db_xref="taxon:10090"
    /clone="D130043B15"
    /clone_1b="RIKEN full-length enriched, 12 days embryo
    spinal ganglion"
    /tissue_type="spinal ganglion"
    /dev_stage="12 days embryo"
    /lab_host="DH10B"
    /note="Site 1: Salt. Site 2: BamHI. cDNA library was
    prepared and sequenced in Mouse Genome Encyclopedia
    Project of Genome Exploration Research Group in Riken
    Genomic Sciences Center and Genome Science Laboratory in
    RIKEN. Division of Experimental Animal Research in Riken
    contributed to prepare mouse tissues. 1st strand cDNA was
    primed with a primer [5',
    GAGAGAGAGCGCGCGCACTGAGTTTCTTTTCTTTTNN 3']. cDNA was
    prepared by using trehalose thermo-activated reverse
    transcriptase and subsequently enriched for full-length by
    cap-trapper. Second strand cDNA was prepared with the
    primer adapter of sequence [5',
    GAGAGAGAGATCTCGAGTTAATTAATTAATTCACCCCGCCCGCC 3']. cDNA
    was cleaved with BamHI and XhoI. Vector: a modified
    pBluescript KS(+) after bulk excision from Lambda FLX I."
BASE COUNT 178 a 180 c 169 g 122 t
ORIGIN
Query Match 17.4%; Score 633.6; DB 10; Length 650;
Best Local Similarity 99.2%; Pred. No. 2.4e-144;
Matches 636; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
      2205  GGAGAGTGAATAAGTGTATATTAACGAGCGCTGTACCCCTTTGAATTCAGAACTACGA 2264
      1  GGAGAGTGAATAAGTGTATATTAACGAGCGCTGTACCCCTTTGAATTCAGAACTACGA 60
      2265  TGAGATCAACCTCCAGCGAGAGATATGATGATGATGAAGCCAGACTGAGAGGCC 2324
      61  TGAGATCAACCTCCAGCGAGAGATATGATGATGATGAAGCCAGACTGAGAGGCC 120
      2325  AGGATGCGTTGGAGAGAGCTGAAGAGGAGACGGATGTTTCCCTGCAACTATGACGA 2384
      121  AGGATGCGTTGGAGAGAGCTGAAGAGGAGACGGATGTTTCCCTGCAACTATGACGA 180
      2385  AAAGATTCGAGAAATGAGTTTCCCACTCCAGCCAAACCACTGACATGATGCG 2444
      181  AAAGATTCGAGAAATGAGTTTCCCACTCCAGCCAAACCACTGACATGATGCG 240
      2445  CCGTCCCGCAAACTGGCTGCGGAGAGCCCGTCTTGGCAGTGACCTTTCTGA 2504

```

Db 241 CCCTGCCCCCAACTGGCTCTGCTAGAGACCCCTGCTCTTTGCCAGTACCTCTTTCTGA 300  
 QY 2505 GCCCTTCACAAACCCCAACAACTGGGAGAGCTTCACTTCCAGCTGGCCCAAGAGCTCAAA 2564  
 Db 301 GCCCTTCACAAACCCCAACAACTGGGAGAGCTTCACTTCCAGCTGGCCCAAGAGCTCAAA 360  
 QY 2565 CGAAGAACCCCAAGAGAGCACTGGGAGAGCTTCACTTCCAGCTGGCCCAAGAGCTCAAA 2624  
 Db 361 CGAAGAACCCCAAGAGAGCACTGGGAGAGCTTCACTTCCAGCTGGCCCAAGAGCTCAAA 420  
 QY 2625 TAGTGTGGGCACTTACGGAGAGATCAGCTTACCCAGCAGCAGCAGCTGGCTCTC 2684  
 Db 421 TAGTGTGGGCACTTACGGAGAGATCAGCTTACCCAGCAGCAGCAGCTGGCTCTC 480  
 QY 2685 CCCATCTCCCTGCTGCTGGGAGAGAGTGAAGAGGAGTCAAGAGCAGCAGCTGTA 2744  
 Db 481 CCCATCTCCCTGCTGCTGGGAGAGAGTGAAGAGGAGTCAAGAGCAGCAGCTGTA 540  
 QY 2745 TCCCTGGAGAGCCCAAAAGACCACTTAATTTTAAACAAAGTGACGTATCACTG 2804  
 Db 541 TCCCTGGAGAGCCCAAAAGACCACTTAATTTTAAACAAAGTGACGTATCACTG 600  
 QY 2805 TCTGGAACAGCAAGACATGCTGCTTGGAGAGAGTTCAG 2845  
 Db 601 TCTGGAACAGCAAGACATGCTGCTTGGAGAGAGTTCAG 641

RESULT 9 610 bp mRNA linear EST 29-MAY-2002  
 LOCUS B0443095  
 DEFINITION UI-M-EVO-bxf-j-12-0-UI-r1 NIH\_MBP\_EVO Mus musculus cDNA clone  
 ACCESSION B0443095  
 VERSION B0443095  
 KEYWORDS B0443095.1 GI:21246207  
 SOURCE EST  
 ORGANISM house mouse.  
 Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 610)  
 NIH-MGC http://mgs.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgaab@imail.nih.gov  
 Tissue Procurement: Dr. James Lin, University of Iowa  
 cDNA library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/BLN at:  
 http://image.llnl.gov  
 This clone was contributed by the Brain Molecular Anatomy Project  
 (BMAP)  
 Seq primer: pYX-5.  
 Location/Qualifiers  
 1. 610  
 /organism="Mus musculus"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5707499"  
 /clone\_1ib="NIH BMAP EVO"  
 /tissue\_type="whole brain"  
 /dev\_stage="embryo 15.5 dpc"  
 /lab\_host="DH10B (T1 phage resistant)"  
 /note="Organ: brain; Vector: pYX-asc; Site: 1. Ecor I;  
 Site 2: Not I; The library was constructed according to  
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. Denatured mRNA was size fractionated on a 1% agarose  
 gel. First strand cDNA synthesis was primed with an  
 oligo-dT primer containing a Not I site. Double stranded  
 cDNA was size selected according to mRNA size fraction,  
 ligated with Ecor I adaptor, digested with Not I, and then

cloned directionally into pYX-asc vector. The library tag  
 sequence located between the Not I site and the polyA tail  
 , is GTGGGTGGA. This library was created for the  
 University of Iowa Mouse Brain Molecular Anatomy Project  
 (BMAP): 'Gene Discovery in the Developing Mouse Nervous  
 System', supported by National Institutes of Mental Health  
 (NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT 175 a 147 c 165 g 123 t  
 ORIGIN  
 Query Match 16.7%; Score 606.8; DB 14; Length 610;  
 Best Local Similarity 99.7%; Pred. No. 8.7e-138;  
 Matches 608; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 2899 AGAATTCACAAACATCGATCTAGTGGCCCTACTGAAAGCTCTGACTTAAGAGAGTG 2958  
 Db 1 AGAATTCACAAACATCGATCTAGTGGCCCTACTGAAAGCTCTGACTTAAGAGAGTG 60  
 QY 2959 GCTTCCCGGCGCCAAAGCCATTCCTCGAGAGAGATTATGTCATGACATAC 3018  
 Db 61 GCTTCCCGGCGCCAAAGCCATTCCTCGAGAGAGATTATGTCATGACATAC 120  
 QY 3019 GAGAGTCTGAGCAAGAGATTAACTTTCAAGCAAGGAGATGATGTTAGCCAAAG 3078  
 Db 121 GAGAGTCTGAGCAAGAGATTAACTTTCAAGCAAGGAGATGATGTTAGCCAAAG 180  
 QY 3079 AAAGATGATGATCTGATGAGAGCGGAAACGATGCGGAGCTTCCCTTTTAC 3138  
 Db 181 AAAGATGATGATCTGATGAGAGCGGAAACGATGCGGAGCTTCCCTTTTAC 240  
 QY 3139 TATGTAGGCTTAAAGATTCAAGAGGCTCTGAACTGCTGGGAAACAGAGAGTTAGGA 3198  
 Db 241 TATGTAGGCTTAAAGATTCAAGAGGCTCTGAACTGCTGGGAAACAGAGAGTTAGGA 300  
 QY 3199 AAAAACCCTGAATTTGCCCGGCTTATGCTTCTTCAAGCTCTGCTCCGAAACAATC 3258  
 Db 301 AAAAACCCTGAATTTGCCCGGCTTATGCTTCTTCAAGCTCTGCTCCGAAACAATC 360  
 QY 3259 ACCCTGCTCTGCGGAGCTGATTTCTGATCCGAAAGAAACCAAGATGATGGTGA 3318  
 Db 361 ACCCTGCTCTGCGGAGCTGATTTCTGATCCGAAAGAAACCAAGATGATGGTGA 420  
 QY 3319 GGAGAACTGCAAGCTCGAGGAGAAACCGCCAGTAGGGTGTTCAGAAATATATGTC 3378  
 Db 421 GGAGAACTGCAAGCTCGAGGAGAAACCGCCAGTAGGGTGTTCAGAAATATATGTC 480  
 QY 3379 AAATCTTAACCCCGGAAACCAAGCAAAATCACCCCACTGAGCTACCAAGACCGAGTG 3438  
 Db 481 AAATCTTAACCCCGGAAACCAAGCAAAATCACCCCACTGAGCTACCAAGACCGAGTG 540  
 QY 3439 CAGCAGAGAGTGTCCAGAGTGTATCGGATGTAAGTACACCCCGAGAAAGATGACGA 3498  
 Db 541 CAGCAGAGAGTGTCCAGAGTGTATCGGATGTAAGTACACCCCGAGAAAGATGACGA 600  
 QY 3499 CTAGCCTTCA 3508  
 Db 601 CTAGCCTTCA 610  
 RESULT 10 603 bp mRNA linear EST 15-MAR-2000  
 LOCUS A1594919  
 DEFINITION m179g10.y1 Soares mouse p3NM19.5 Mus musculus cDNA clone  
 IMAGE:482370 5' similar to TR:042287 INTERSECTIN., mRNA  
 sequence.  
 ACCESSION A1594919  
 VERSION A1594919.1 GI:4603967  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 603)







```

Db      8  CGCTTCCCCCTGATGAAACAGCAACCGTGTCTATTTCATGTGACACGACATTTGGTA 67
Qy      350 TAGAGAGGATTTGCTAGCATGCCACCACTACAGCTTTGCTCTGCAATGGGCTCCA 409
Db      68  TAGAGAGGATTTGCTAGCATGCCACCACTACAGCTTTGCTCTGCAATGGGCTCCA 127
Qy      410 TTCCAGTTGTTGAATGTCTCCACCCCTTAGTATCTTCTGCTCCAGACAGAGTGCCTC 469
Db      128  TTCCAGTTGTTGAATGTCTCCACCCCTTAGTATCTTCTGCTCCAGACAGAGTGCCTC 187
Qy      470 CCTGCTAACGGGGGCTCCCTCCGTCATACAGCTGCTGCTTTGGCAATCCTGCGAG 529
Db      188  CCTGCTAACGGGGGCTCCCTCCGTCATACAGCTGCTGCTTTGGCAATCCTGCGAG 247
Qy      530 CCACATGCGCAAGAGTTCTTCTCTCAGACAGATGTGTCAGAGGTACAAATTAACACTA 589
Db      248  CCACATGCGCAAGAGTTCTTCTCTCAGACAGATGTGTCAGAGGTACAAATTAACACTA 307
Qy      590 AGTTACAGAAAGGACATCATTTGATGTGCGCAGCGCCCTCCAGACAGAGATGGGCTG 649
Db      308  AGTTACAGAAAGGACATCATTTGATGTGCGCAGCGCCCTCCAGACAGAGATGGGCTG 367
Qy      650 TGCTCTAGTCATTAAGGCTGAATTAACAGGAGTATTCAACAGCCAGCAAAACTATGA 709
Db      368  TGCTCTAGTCATTAAGGCTGAATTAACAGGAGTATTCAACAGCCAGCAAAACTATGA 427
Qy      710 GTGACACATTAACAGGTCGCCAGCAAGAACTATTCTCATGCAATCAAGTTTACCCTGAG 769
Db      428  GTGACACATTAACAGGTCGCCAGCAAGAACTATTCTCATGCAATCAAGTTTACCCTGAG 487
Qy      770 CTGAGCTGGCTTCAATATGGAATCTTTCTGACATTTGATCAAGATGGAATCTCATGCGAG 829
Db      488  CTGAGCTGGCTTCAATATGGAATCTTTCTGACATTTGATCAAGATGGAATCTCATGCGAG 547
Qy      830 AAGAAATTTATCTTCACTATGACCTTAATGATGTGTCATGCTGCTGCTGCAACCACTGCCG 888
Db      548  AAGAAATTTATCTTCACTATGACCTTAATGATGTGTCATGCTGCTGCTGCAACCACTGCCG 606

RESULT 12
BM681943/c 773 bp mRNA linear EST 27-FEB-2002
LOCUS      UI-E-EO1-aiw-e-23-0-UI-s1 UI-E-EO1 Homo sapiens cDNA clone
DEFINITION UI-E-EO1-aiw-e-23-0-UI 3', mRNA sequence.
ACCESSION  BM681943
VERSION     BM681943.1 GI:1891839
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 773)
            Normalization and subtraction: two approaches to facilitate gene
            discovery
            Genome Res. 6 (9), 791-806 (1996)
            97044477
JOURNAL    Contact: Soares, MB
MEDLINE    Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.iowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA
sequence: 1-53, >Poly_A$imple_repeat (matched complement)

```

```

FEATURES
source
Seq primer: M13 Forward
POLYA=Yes.
Location/Qualifiers
1..773
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-EO1-aiw-e-23-0-UI"
/tissue_type="fetal eye"
/dev_stage="fetal"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pT73-Pac (pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-E-EO1 is a normalized cDNA library containing the
following tissue(s): fetal eye. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT73-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
CGCGTATACC. This library was created for the program, Gene
Discovery in the Visual System, supported by National Eye
Institute (NEI).
TAG_LIB=UI-E-EO1
TAG_TISSUE=human fetal eye
TAG_SEQ=CGCGTATACC"

BASE COUNT  118 a 182 c 130 g 343 t

ORIGIN
Query Match      15.5%; Score 563.4; DB 14; Length 773;
Best Local Similarity 83.1%; Pred. No. 4,2e-127;
Matches 642; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

Qy      1168 CAGGAGAGAAAGACCGGAGCGCCAGAGAGAGAGCCAAAGCGGAGCTGAGCTGGAG 1227
Db      773  CAGAGAGGAAAGAGCGGTGAGCGCCAGAGAGCAAGGCGCAAAAGACAACTGGAATCTGGAG 714
Qy      1228 AAGCACTGAGAAAGACCGGAGCTGAGCGGCGAGCGAGAGAGAGAGAGAGAGAG 1287
Db      713  AAGCACTGAGAAAGACCGGAGCTGAAACGCGAGAGAGAGAGAGAGAGAGAGAGAG 654
Qy      1288 ATCCAGAGCGCGGAGCGCGGCAAAACGGGAACCTGGAAGGACGACCACTGATGGGAA 1347
Db      653  ATTGAGAGCGGAGAGGCTGCAAAACGGGAACCTGGAAGGACGACCACTGATGGGAA 594
Qy      1348 CGAAACCGGAGACGAAACTCTGAAATCAGAGCAACAGAGACGAGGACCGTGGTC 1407
Db      593  CGGAATCGAAGCAAGAACTTAATCAAGAAACAAAGACAGAGACATAGTTGTA 534
Qy      1408 CTGAAGGACAGAGAGAGACTCTGGAAGTTGAGTTAGAGCTCTGAATGACAAAAGCAT 1467
Db      533  CTGAAGGACAAAGAAAGAGACTTTGGAATTTGAATTAAGACTCTTAATGATTAAGAAACAT 474
Qy      1468 CAGCTAGAAGGAAACTCTGAGATATCAGGTGCTGACCAACCCAGAGCAAGAAATT 1527
Db      473  CACTAGAAGGAAACTCTCAAGATATCAGATGTCAGATTCACACCCAAAGCAAGAAATT 414
Qy      1528 GAGAGACGAAACAAGTCTAGAGACTTAAGAAATTGCTGAATTCACCCACTTACAGACAG 1587
Db      413  GAGAGACCAAAACAATCTAGAGATTGAGAAATTCGCAATTCACCCACTTACAGACAA 354
Qy      1588 TTGACGAAATCTCAGCAAAATGCTTGAAGACTATTTCAGAGAAACGATATCTGATGAC 1647
Db      353  TTACAGGAATCTCAGCAAAATGCTTGAAGACTATTTCAGAGAAACGATATCTGATGAC 294
Qy      1648 CAGTTAAACAAAGTCCAGAGAAACAGTTTGATATGAGACTCGCTTCTTACCCTCAAGA 1707
Db      293  CAATTAACAAAGTTCCAGAGAAACAGTTTGATATGAGACTTCTTGTACATTAAGA 234

```

```

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-DQ1-cl1-21-0-U"
/clone_lib="UI-R-DQ1"
/seq_features="adult"
/lab_host="DH10B (Life Technologies)"
/notes="Vector: pT73-Pac (Pharmacia) with a modified
polylinker site. I: Not I; Site 2: Eco RI; The UI-R-DQ1
library is a normalized Rat Cell Line R337-5A library
(mRA) constructed in pT73 PAC vector according to the
procedure described by Bonaldo, Lennon & Soares
(Normalization and Subtraction: Two Approaches to
Facilitate Gene Discovery. Genome Research 6: 791-806,
1996). The oligonucleotide used to prime first strand
synthesis contained the sequence tag GCACGACGAC between
the Not I cloning site and d18 stretch. The Rat Cell Line
R337-5A was provided by Mary Hendrix of the University of

```

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 1062)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.

Email: cgarbs-r@mail.nih.gov  
 Tissue Procurement: Dr. James R. Lupski  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 Plate: LLM13562 row: n column: 16  
 High quality sequence start: 116  
 High quality sequence stop: 760.  
 Location/Qualifiers

## FEATURES

1..1062  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:6180303"  
 /clone\_lib="Lupski\_dorsal\_root\_ganglion"  
 /sex="male"  
 /tissue\_type="dorsal root ganglia"  
 /dev\_stage="adult, 36 yr"  
 /lab\_host="DH10B"  
 /note="Vector: PCMV-SPORT6 (Life Technologies); Site\_1:  
 NotI; Site\_2: SalI; cDNA made by oligo-dT priming.  
 Directionally cloned using the following adaptors:  
 5'-TGACCCAGCGCTCG-3' and  
 5'-GACTAGTCTAGATCGGAGCGCGCCCT(15)-3'. Size selected >  
 1 kb for average insert length 1.7 kb. This is a primary  
 library, non-amplified. Library constructed by Life  
 Technologies and donated by J. Lupski, M.D./Ph.D (Baylor  
 College of Medicine) and is available through Life  
 Technologies."

BASE COUNT 338 a 250 c 261 g 209 t 4 others  
 ORIGIN

Query Match 15.2%; Score 555; DB 14; Length 1062;  
 Best Local Similarity 90.1%; Pred. No. 5.5e-125;

Matches 594; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

OY 2984 TTCGCCGAGAGAGATTATTCATGTACACATACAGAGTTCTGAGCAAGAGATTAA 3043  
 DB 113 TTCCGGGAGAGAGATTATTCATGTACACATACAGAGTTCTGAGCAAGAGATTAA 172  
 OY 3044 CCTTCAGCAAGGGAGTGTATTTGTTACCAAGAAAGTGTGTGACGCGGAA 3103  
 DB 173 CCTTCAGCAAGGGAGTGTATTTGTTACCAAGAAAGTGTGTGACGCGGAA 232  
 OY 3104 CGGTGGCCGCAAGTCCGAGTCTTCCCTTTACTATGTAGAGCTTAAAGATTCAAGG 3163  
 DB 233 CAGTGGGCGCAAGGCGGAGTCTTCCCTTTACTATGTAGAGCTTAAAGATTCAAGG 292  
 OY 3164 GCTCTGAACTGCTGGGAAACAGGAGTTTAGGAAAAAACCCTGAATTGCCAGGTTA 3223  
 DB 293 GCTCTGAACTGCTGGGAAACAGGAGTTTAGGAAAAAACCCTGAATTGCCAGGTTA 352  
 OY 3224 TTGCTTCTACGCTGTACTGTGTCCGAAACAACCTGCTGCTGGGCAAGTGTATTC 3283  
 DB 353 TTGCTTCTACGCTGTACTGTGTCCGAAACAACCTGCTGCTGGGCAAGTGTATTC 412  
 OY 3284 TGATCCGAAAAAAGAACCCAGGTGTGTGGGAGAGAACTGCAAGTCCAGGGAAAA 3343  
 DB 413 TGATCCGAAAAAAGAACCCAGGTGTGTGGGAGAGAACTGCAAGTCCAGGGAAAA 472  
 OY 3344 AGGCGAGATAGGGGTGTTCCAGCAAAATTATGTCAAACTTCAACCCCGGAAACAAGA 3403  
 DB 473 AGGCGAGATAGGGGTGTTCCAGCAAAATTATGTCAAACTTCAACCCCGGAAACAAGA 532  
 OY 3404 AAATCAACCCCACTGAGTACCCAGACCGCAGTGCAGCAGCAGTGTCCAGGTATCG 3463  
 DB 533 AAATCACTCCAAACAGAGCCACTAAGTCAACAGACATTAGGGGAGTGTCCAGGTATCG 592  
 OY 3464 GATGTATGATTAACCGCCCGAGAGAGATGACAACTAGCTTCAGCAAAAGGCGAGATCA 3523  
 DB 593 GATGTATGATTAACCGCCCGAGAGAGATGACAACTAGCTTCAGCAAAAGGCGAGATCA 652

OY 3524 TCAACGTCCTCAACAGAGAGAGCCGAGCTGTGTGAAAAAGAGAACTCAGTGGCAAGTTG 3583  
 DB 653 TCAACGTCCTCAACAGAGAGAGCCCTCAGTGTGTGAAAAAGAGAACTCAGTGGCAAGTTG 712  
 OY 3584 GGCCTTCCCACTCAATTAATTAAGCTGACACAGACATGAGACCCGAGCAGCAATGA 3642  
 DB 713 GGCCTTCCCACTCAATTAATTAAGCTGACACAGACATGAGACCCGAGCAGCAATGA 771

## RESULT 15

## LOCUS

BE373101 724 bp mRNA linear EST 21-JUL-2000  
 DEFINITION 601224527F1 NCI\_CGAP\_Mam1 Mus musculus cDNA clone IMAGE:3582726 5',  
 mRNA sequence.

## ACCESSION

BE373101  
 VERSION BE373101.1 GI:9318464

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

NIH-MGC <http://mgc.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgarbs-r@mail.nih.gov  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 Plate: LLM6738 row: f column: 07  
 High quality sequence stop: 571.  
 Location/Qualifiers

## FEATURES

## source

1..724  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:3582726"  
 /clone\_lib="NCI CGAP Mam1"  
 /tissue\_type="tumor, biopsy sample"  
 /dev\_stage="10 months, virgin"  
 /lab\_host="DH10B"  
 /note="Organ: mammary; Vector: PCMV-SPORT6; Site\_1: SalI;  
 Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
 Library constructed by Life Technologies. Investigator  
 providing samples: Gilbert Smith, NIH"

## BASE COUNT

181 a 185 c 245 g 113 t

## Query Match

14.8%; Score 537.8; DB 10; Length 724;  
 Best Local Similarity 97.9%; Pred. No. 7.8e-121;

Matches 608; Conservative 0; Mismatches 7; Indels 6; Gaps 6;

OY 695 ACGAACAACCTATAGTGTGACACTTAAACAGTCCCGGCAAGAACTATCTCATGCAAT 754  
 DB 1 ACGAACAACCTATAGTGTGACACTTAAACAGTCCCGGCAAGAACTATCTCATGCAAT 60  
 OY 755 CAAGTTAACCCAGAGCTCAGTGGCTTCAATATGAGATCTTTGACATGATGATCAAGATG 814  
 DB 61 CAAGTTAACCCAGAGCTCAGTGGCTTCAATATGAGATCTTTGACATGATGATGATG 119  
 OY 815 GAAATCACTAGTCAAGAAATTTATCTAGTATGACACTTAATGATGATGATGATGATG 874  
 DB 120 GAAATCACTAGTCAAGAAATTTATCTAGTATGACACTTAATGATGATGATGATGATG 179  
 OY 875 GTACAGCACTGCGCCGCTGCTGCTTCAAGATATCATCTCTCTTCAAGAGATTTC 934  
 DB 180 GTACAGCACTGCGCCGCTGCTGCTTCAAGATATCATCTCTCTTCAAGAGATTTC 239

QY 935 GCTCCGCGAGTGGGATGTCCGTACATAAGCTCTTCTTGTGATCAGAGGCTGCTGAG 994  
 Db 240 GCTCCGCGAGTGGGATGTCCGTACATAAGCTCTTCTTGTGATCAGAGGCTGCTGAG 299  
 QY 995 AGCGGTGTCAGAGATGAGCAGCAGCCAGAGAACTTGCTGTGACATTTGAAGATA 1054  
 Db 300 AGCGGTGTCAGAGATGAGCAGCAGCCAGAGAACTTGCTGTGACATTTGAAGATA 357  
 QY 1055 AGAAGCGGGAACCTTGAGGGAGGAGGTGAGCTGAGAAAGCGCCGCAAGCGCTCT 1114  
 Db 358 AGAAGCGGGAACCTTGAGGGAGGAGGTGAGCTGAGAAAGCGCCGCAAGCGCTCT 416  
 QY 1115 TGGAGCAGCAGCGCAAGAGAGAGAGCGGTGCTCAGCTGAGCGCGCCGAGCAGAGA 1174  
 Db 417 TGGAGCAGCAGCGCAAGAGAGAGAGCGGTGCTCAGCTGAGCGCGCCGAGCAGAGA 476  
 QY 1175 GGAAGAGCGGAGCGCCAGAGAGAGAG--GCCAAGCGGCACTGAGCTGAGAAAGCAG 1233  
 Db 477 GGAAGAGCGGAGCGCCAGAGAGAGAGAGCAATGCGGCACTGAGAGCTGAGAAAGCAG 536  
 QY 1234 CTGAGAGAGCAGCGGAGCTGAGCGGAGAGAGAGAGAGAGAGAGAGAGATGAG 1293  
 Db 537 CTGAGAGAGCAGCGGAGCTGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG--GATGAG 595  
 QY 1294 AGGCGGAGAGCGCCAAAGCG 1314  
 Db 596 ACGCGGAGAGCGCCAAAGAGG 616

Search completed: March 14, 2003, 11:05:44  
 Job time : 4408.26 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Comphen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2003, 02:10:41 ; Search time 69 Seconds  
(without alignments)  
2342.507 Million cell updates/sec

Title: US-09-674-237A-3

Perfect score: 6263 1 MAQPTPFGSLDVMATIVE.....VGLFSPSNVYKLTMDPSPQ 1213

Sequence: BLOSUM62

Scoring table: Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_101002:\*

1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*

2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*

3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*

4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*

5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*

6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*

7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*

8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*

9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*

10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*

11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*

12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*

13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*

14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*

15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*

16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*

17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*

18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*

19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*

20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*

21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*

22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*

23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6258.5	99.8	1214	21	AAV57444
2	6258.5	99.8	1715	21	AAV57449
3	5860	93.5	1215	20	AAV32156
4	5847.5	93.3	1220	20	AAV32155
5	5442.5	86.8	1144	20	AAV32154
6	4277	68.2	1031	22	AAU87168
7	4277	68.2	1035	22	AAU43519
8	3097.5	49.4	932	22	ABG20575
9	3091.5	49.3	1681	22	AAU17881
10	3070.5	49.0	648	20	AAV32157

11	3068	48.9	1697	22	AAU79199	Human protein SEQ
12	3049.5	48.6	1197	21	AAV57445	Mouse Ees2 protein
13	3049.5	48.6	1658	21	AAV57450	Mouse Ees2L protei
14	3046	48.6	1683	21	AAV71160	Rat phosphodiester
15	3034.5	48.4	642	22	AAW63229	Human polypeptide
16	3012	48.0	641	20	AAV32158	Human SH3D1A prote
17	2812	44.9	595	22	AAU84691	Human protein sequ
18	2273.5	36.3	471	22	AAW63676	Human polypeptide,
19	2058.5	32.8	462	17	AAW05395	Human SH3P17 prote
20	1702	27.1	1097	22	ABU57814	Drosophila melanog
21	1536.5	24.5	936	22	ABG20576	Novel human diagno
22	1388.5	22.1	509	17	AAW05399	Human clone 65 pro
23	1361	21.7	990	22	AAU87256	Human central nerv
24	1330.5	21.2	530	22	AAU17173	Novel signal trans
25	1183.5	18.9	436	22	AAU83738	Human protein sequ
26	816.5	13.0	248	17	AAW05396	Human SH3P18 prote
27	774.5	12.4	182	22	AAU00069	Human polypeptide
28	712	11.4	164	23	ABU41930	Human ovarian anti
29	572	9.1	144	22	AAU87484	Novel central nerv
30	572	9.1	144	22	AAU3613	Human polypeptide
31	572	9.1	144	22	AAU19936	Human polypeptide
32	536.5	8.6	866	21	AAU6194	Nuclear transport
33	535.5	8.5	864	22	AAU0292	Human polypeptide
34	527	8.0	813	23	AAU02924	Protein of App rel
35	503	8.0	754	22	AAU85697	Human protein sequ
36	486.5	7.8	896	17	AAU82750	Human EGF receptor
37	486.5	7.8	896	19	AAU47117	Human eps15 protei
38	486.5	7.8	896	20	AAU94405	Human eps15 protei
39	486	7.8	897	17	AAU82751	Murine EGF recepto
40	486	7.8	897	19	AAU47118	Murine eps15 prote
41	486	7.8	897	20	AAU94406	Drosophila melano
42	485	7.7	668	22	ABU59618	Human central nerv
43	431.5	6.9	224	22	AAU87171	Human polypeptide
44	431.5	6.9	224	22	AAU43520	Novel human calciu
45	431.5	6.9	224	22	AAU19935	Novel human calciu

#### ALIGNMENTS

RESULT 1

AAV57444

ID AAV57444 standard; Protein; 1214 AA.

XX

AC AAV57444;

XX

DT 28-FEB-2000 (first entry)

XX

XX

DE Mouse Ees1 protein sequence.

XX

XX

DE Mouse; murine; Ees1; Ees2; endocytosis; vesicular trafficking;

KW regulation; actin cytoskeleton; detection; cancer; infection;

KW EH-domain and SH3-domain regulator of endocytosis; anticancer;

KW antiproliferative; antiviral.

XX

XX

OS Mus sp.

XX

XX

PN W0955728-A2.

XX

PD 04-NOV-1999.

XX

XX

PF 27-APR-1999; 99WO-CA00375.

XX

PR 27-APR-1998; 98CA-2230201.

XX

PR 05-FEB-1999; 99US-0118739.

XX

PA (HSCR-) HSC RES & DEV LP.

XX

PI Egan SE, Wang W, Sengar A;

XX

XX

DR WPI; 2000-052802/04.

XX

DR N-PSDB; AAU39008; AAU39009.

XX

PT New nucleic acid encoding Ees1 and 2 proteins, involved in regulation  
 PT of endocytosis, used e.g. for treating cancer or preventing viral  
 PT infection  
 XX  
 PS Claim 14; Page 43; 99pp; English.  
 CC The present sequence represents mouse Ees1. The present invention  
 CC specifically describes mammalian Ees1 and 2 proteins (I) and their splice  
 CC variants (Ese = EH-domain and SH3-domain regulator of endocytosis). (I)  
 CC are involved in regulation of clathrin-mediated endocytosis (as a complex  
 CC with Esp1s protein), vesicular trafficking and actin cytoskeleton.  
 CC Generally (I) (or its antagonists, mimetics, fragments and inactive  
 CC mutants); (I)-specific antibodies (Ab); sequences antisense to the (I)  
 CC polynucleotide; agents that downregulate expression of Ees genes or  
 CC antagonists of an Ese binding partner are used to treat diseases  
 CC associated with undesirable endocytosis and resulting changes in cellular  
 CC function. Particularly overexpression of Ees1 is used to block  
 CC clathrin-mediated endocytosis in vivo or in cell cultures, while  
 CC administration of (I) is used to promote endocytosis of selected cells.  
 CC (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of  
 CC cells that can be stimulated to proliferate by a growth factor receptor;  
 CC and similar compounds (also inactive Ese mutants) can be used to prevent  
 CC viral infection. Endocytosis may also be regulated, in vivo or in cell  
 CC cultures, by forming an Ese-Esp1s complex, then binding dynamin to the  
 CC complex. Generally conditions that can be treated include cancer;  
 CC abnormal cell division or migration that can be treated include cancer;  
 CC receptor signalling, tissue development or synaptic transmission.  
 XX  
 SQ Sequence 1214 AA:  
 Query Match 99.8%; Score 6258.5; DB 21; Length 1214;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1213; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 600 DFNNOULKELEIHSKQOOLQKORSTFAARLUKOKROEKSLLELEKOKEDAPRVQEBDKOM 659  
 Db 601 DFNNOULKELEIHSKQOOLQKORSTFAARLUKOKROEKSLLELEKOKEDAPRVQEBDKOM 660  
 QY 660 LEHVQOEOPRRPKPHEDRLKREDSYRKKEAEERAKPEMOKOSRLFHPHOEPAKLATQ 719  
 Db 661 LEHVQOEOPRRPKPHEDRLKREDSYRKKEAEERAKPEMOKOSRLFHPHOEPAKLATQ 720  
 QY 720 APWSTTEGPIITIAQOSVKKVYRRLALYPFESRSHDITITOPGDIVWVDSQGEPEWLG 779  
 Db 721 APWSTTEGPIITIAQOSVKKVYRRLALYPFESRSHDITITOPGDIVWVDSQGEPEWLG 780  
 QY 780 GELGKGTWFPANYAEKIPENEVPTPAKPYTDLTSAAPKALARETPAPUPVTSSEPTT 839  
 Db 781 GELGKGTWFPANYAEKIPENEVPTPAKPYTDLTSAAPKALARETPAPUPVTSSEPTT 840  
 QY 840 PNNMADFSTWSSSNEKETDNDWTAOPLTVPASAGLROKRSATPATATGSSPSPV 899  
 Db 841 PNNMADFSTWSSSNEKETDNDWTAOPLTVPASAGLROKRSATPATATGSSPSPV 900  
 QY 900 LGQGEKVEGLQAQALYPRAKDNHLPNKSVDITVLEQDDMMWFGEVQOKGWFPSYV 959  
 Db 901 LGQGEKVEGLQAQALYPRAKDNHLPNKSVDITVLEQDDMMWFGEVQOKGWFPSYV 960  
 QY 960 KLISGPVKSTSIDTGPESPASLKRVASPAKPAIPGEBFIAMTYESSEBOGDLTPQOG 1019  
 Db 961 KLISGPVKSTSIDTGPESPASLKRVASPAKPAIPGEBFIAMTYESSEBOGDLTPQOG 1020  
 QY 1020 DVIIVTKKDGDMWTGTVGDKSGVFPSSNVYRLKDSGSGTGKTKSLGKKEPIAOVIA5YA 1079  
 Db 1021 DVIIVTKKDGDMWTGTVGDKSGVFPSSNVYRLKDSGSGTGKTKSLGKKEPIAOVIA5YA 1080  
 QY 1080 ATGPEQLTLAPGOLILIRKKNPGGMEGELQARGKKQIGFPPANYVKLSPGSKITPT 1139  
 Db 1081 ATGPEQLTLAPGOLILIRKKNPGGMEGELQARGKKQIGFPPANYVKLSPGSKITPT 1140  
 QY 1140 ELFKTAVQPAVCQVIGMYDYTAQNDDELAIFSGQIINVINKEDPDMWKGVEVGQVGLFPS 1199  
 Db 1141 ELFKTAVQPAVCQVIGMYDYTAQNDDELAIFSGQIINVINKEDPDMWKGVEVGQVGLFPS 1200  
 QY 1200 NYVKLTITMDPSQC 1213  
 Db 1201 NYVKLTITMDPSQC 1214

RESULT 2  
 ID AA57449 standard; Protein: 1715 AA.  
 XX  
 XX AA57449;  
 DT 28-FEB-2000 (first entry)  
 XX  
 DE Mouse Ees1 protein sequence.  
 XX  
 KW Mouse; murine; Ees1, Ees2; endocytosis; vesicular trafficking;  
 KW regulation; actin cytoskeleton; detection; cancer; infection;  
 KW EH-domain and SH3-domain regulator of endocytosis; anticancer;  
 KW antiproliferative; antiviral.  
 XX  
 OS Mus sp.  
 XX  
 PN W09955728-A2.  
 XX  
 PD 04-NOV-1999.  
 XX  
 XX 27-APR-1999; 99WO-CA00375.  
 PF  
 XX 27-APR-1998; 98CA-2230201.  
 PR 05-FEB-1999; 99US-0118739.  
 XX  
 PA (HSCR-) HSC RES & DEV LP.

XX Egan SE, Wang W, Sengar A;  
 XX MPI: 2000-052802/04.  
 DR N-PSDB; AA339024, AA339025.  
 XX  
 PT New nucleic acid encoding Ese1 and 2 proteins, involved in regulation  
 PT of endocytosis, used e.g. for treating cancer or preventing viral  
 PT infection  
 XX  
 PS Claim 14; Page 62-63; 99pp; English.  
 XX  
 CC The present invention specifically describes mammalian Ese1 and 2  
 CC proteins (1) and their splice variants (Ese = EH-domain and SH3-domain  
 CC regulator of endocytosis). (1) are involved in regulation of clathrin-  
 CC mediated endocytosis (as a complex with Esp15 protein), vesicular  
 CC trafficking and actin cytoskeleton. Generally (1) (or its antagonists,  
 CC mimetics, fragments and inactive mutants); (1)-specific antibodies (Ab);  
 CC sequences antisense to the (1) polynucleotide; agents that downregulate  
 CC expression of Ese genes or antagonists of an Ese binding partner are  
 CC used to treat diseases associated with undesirable endocytosis and  
 CC resulting changes in cellular function. Particularly overexpression of  
 CC Ese1 is used to block clathrin-mediated endocytosis in vivo or in cell  
 CC cultures, while administration of (1) is used to promote endocytosis of  
 CC selected cells. (Ant)agonists of (1) or Ab are used to suppress abnormal  
 CC proliferation of cells that can be stimulated to proliferate by a growth  
 CC factor receptor; and similar compounds (also inactive Ese mutants) can be  
 CC used to prevent viral infection. Endocytosis may also be regulated, in  
 CC vivo or in cell cultures, by forming an Ese-Esp15 complex, then binding  
 CC dynamain to the complex. Generally conditions that can be treated include  
 CC cancer; abnormal cell division or migration; viral infection; or abnormal  
 CC receptor signaling, tissue development or synaptic transmission. The  
 CC present sequence represents mouse Ese1 protein sequence.  
 CC  
 XX

SO Sequence 1715 AA;

Query Match 99.8%; Score 6258.5; DB 21; Length 1715;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1213; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 MAQPPPPGSLVMAITVERAKHQOFLSLKPIAGFITGDARNFQSGPQVLAQ 60  
 DB 1 MAQPPPPGSLVMAITVERAKHQOFLSLKPIAGFITGDARNFQSGPQVLAQ 60  
 QY 61 IVALADNNNGRDQVFEFSIAMKLILKLOGYQLPSTLPVMKQGVASAPAFIGGI 120  
 DB 61 IVALADNNNGRDQVFEFSIAMKLILKLOGYQLPSTLPVMKQGVASAPAFIGGI 120  
 QY 121 ASMPPLTAVAPVEMGSIPIVVGMSPLVSSVPRAVPPLANGAPVIOPLPAFAHPAATWP 180  
 DB 121 ASMPPLTAVAPVEMGSIPIVVGMSPLVSSVPRAVPPLANGAPVIOPLPAFAHPAATWP 180  
 QY 181 KSSSFSRSGSGSGLNTKLOKAGSFVVASAPPAEMAVPQSSRLKRYQLNSHDKTMSGHL 240  
 DB 181 KSSSFSRSGSGSGLNTKLOKAGSFVVASAPPAEMAVPQSSRLKRYQLNSHDKTMSGHL 240  
 QY 241 TGPQARTILMOSSLPOKQASINWLSIDIDDGKLTAEFFILAHMLLDVAMSGOPLPVLP 300  
 DB 241 TGPQARTILMOSSLPOKQASINWLSIDIDDGKLTAEFFILAHMLLDVAMSGOPLPVLP 300  
 QY 301 PEYIPPSFRFRVSGSGSVISSSVDOQLPEPSSSEDEQPEKKLPVTFEDKKRENFERG 360  
 DB 301 PEYIPPSFRFRVSGSGSVISSSVDOQLPEPSSSEDEQPEKKLPVTFEDKKRENFERG 360  
 QY 361 SVLEKRRQALLQORKEQERLQLERAERERERQDEAKQOLEKQLEKQELER 420  
 DB 361 SVLEKRRQALLQORKEQERLQLERAERERERQDEAKQOLEKQLEKQELER 420  
 QY 421 QREERKEIEREPAKRELEOROLEWERNROELLNQNKEOGGVVYAKARKTLFEF 480  
 DB 421 QREERKEIEREPAKRELEOROLEWERNROELLNQNKEOGGVVYAKARKTLFEF 480  
 QY 481 LEALNDKQOLEGKLODIRCLATQROEISTNKSRELRIAEITTHLQOOLQESQOMLGR 540

DB 481 LEALNDKQOLEGKLODIRCLATQROEISTNKSRELRIAEITTHLQOOLQESQOMLGR 540  
 QY 541 IPEKOILSDOLKQVQNSLHRDSILTLKRALLEAKELARQOLREOLDEVERETSKLQEID 600  
 DB 541 IPEKOILSDOLKQVQNSLHRDSILTLKRALLEAKELARQOLREOLDEVERETSKLQEID 600  
 QY 601 VFNNQLEKREIHSKQOLQKORSLEAARLKQKEOERKSELEKQEKEDAKORRVERDKQWL 660  
 DB 601 VFNNQLEKREIHSKQOLQKORSLEAARLKQKEOERKSELEKQEKEDAKORRVERDKQWL 660  
 QY 661 EHVOOEEDPRPRKHEEDRLKREDSVRKKEAEPAKEMODKOSRLPHPOEPAKLATQA 720  
 DB 661 EHVOOEEDPRPRKHEEDRLKREDSVRKKEAEPAKEMODKOSRLPHPOEPAKLATQA 720  
 QY 721 PMSTTEKPLTISAQESVKVYVYALYPFESRSHDEITTIQPGDIVMDESQTEGPGWLGG 780  
 DB 721 PMSTTEKPLTISAQESVKVYVYALYPFESRSHDEITTIQPGDIVMDESQTEGPGWLGG 780  
 QY 781 ELKGTGWFPAHYAEKIPENEVPTPAKPVTDLTISAPAPKLARETAPLPVTSSESTTP 840  
 DB 781 ELKGTGWFPAHYAEKIPENEVPTPAKPVTDLTISAPAPKLARETAPLPVTSSESTTP 840  
 QY 841 NNWADFSSTWSSSNEKRETDNMDTMAAOPSLTVPSAGOLRQSAFTPATATGSSPSVL 900  
 DB 841 NNWADFSSTWSSSNEKRETDNMDTMAAOPSLTVPSAGOLRQSAFTPATATGSSPSVL 900  
 QY 901 GQGEVEGLQALYPPRAKDNHNFNKSVDITVLEQODMMFGEVQOGKMPKSYVK 960  
 DB 901 GQGEVEGLQALYPPRAKDNHNFNKSVDITVLEQODMMFGEVQOGKMPKSYVK 960  
 QY 961 LISGVRKSTSIDTGPTESPASLKRVASPAKPAIPGEFPIAMTYESSQGLTFQOGD 1020  
 DB 961 LISGVRKSTSIDTGPTESPASLKRVASPAKPAIPGEFPIAMTYESSQGLTFQOGD 1020  
 QY 1021 VIYVT-KKDGWMTGTGDKSGVFPSSVYVRLKDSGSGTAGKSLGKKPEIAQVIAVSYA 1079  
 DB 1021 VIYVT-KKDGWMTGTGDKSGVFPSSVYVRLKDSGSGTAGKSLGKKPEIAQVIAVSYA 1079  
 QY 1080 ATGPQOLTLAPGOLILIRKKNPGGMEGELQARCKKQIGFPAHYVYKLSPGSKITPT 1139  
 DB 1080 ATGPQOLTLAPGOLILIRKKNPGGMEGELQARCKKQIGFPAHYVYKLSPGSKITPT 1139  
 QY 1140 ELPKTAVQPAVCQYIGWYDYTAQNDDELAFSGQIINVLNKEDPDMKGEVSGVGLFPS 1199  
 DB 1140 ELPKTAVQPAVCQYIGWYDYTAQNDDELAFSGQIINVLNKEDPDMKGEVSGVGLFPS 1199  
 QY 1200 NYVKLTMDPSQ 1213  
 DB 1201 NYVKLTMDPSQ 1214

# RESULT 3

AA32156  
 ID AA32156 standard; Protein; 1215 AA.

AA32156;

DT 01-FEB-2000 (first entry)

DE Human SH3D1A protein.

SH3D1A; human; Down's syndrome; leukemia; cancer;  
 megalocytic abnormality; myeloproliferative disorder;  
 platelet disorder; neural disorder; thrombocytopenia;  
 haematopoietic disorder; cognitive dysfunction; microcephaly;  
 lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy.

OS Homo sapiens.

PN WO9953062-A2.

XX 21-OCT-1999.

PD



XX PF 16-APR-1999; 99WO-US08371.  
 XX PR 16-APR-1998; 98US-0082007.  
 XX PA (CEDA-) CEDARS SINAI HEALTH SYSTEM.  
 XX PI Korenberg JR, Chen X;  
 XX DR WPI: 1999-633829/54.  
 XX PT N-PSDB; AA234572.  
 XX PT Nucleic acid from the human SH3D1A gene and its products, useful for  
 XX PT the diagnosis and treatment of myeloproliferative disorders and  
 XX PT leukaemia -  
 XX PS  
 XX PS Claim 14; Fig 11; 99pp: English.  
 CC This sequence represents the protein encoded by the human SH3D1A  
 CC cDNA clone 11 (see AA234572). SH3D1A contributes to the development  
 CC of platelets and the pathogenesis of leukemias, both in general  
 CC and in particular those involving the megakaryocytic lineage. The  
 CC gene maps to the small candidate region for low platelets on  
 CC chromosome 21. The protein includes SH3 domains and EH domains,  
 CC both associated with protein-protein interactions and the latter  
 CC with maintenance of the cytoskeleton. At least 3 isoforms of  
 CC SH3D1A exist (see AA23154-58). The invention provides methods for  
 CC the diagnosis and treatment of megakaryocytic abnormality,  
 CC myeloproliferative disorder, platelet disorder, acute leukaemia,  
 CC neural disorders, thrombocytopenia, platelet disorder on  
 CC chromosome 21, low platelets in deletion for 21, association of  
 CC gains in chromosome 21 with leukemias, neural abnormalities,  
 CC dysfunctions and disorders including brain malformations and  
 CC corresponding cognitive dysfunctions, microcephaly, lissencephaly,  
 CC and colpocephaly. Methods are also provided: for suppressing  
 CC cells unable to regulate themselves; screening for a somatic  
 CC alteration in the SH3D1A gene; monitoring the progress and  
 CC adequacy of a treatment; monitoring tumour risk progress or  
 CC megakaryocytic abnormality, myeloproliferative disorder,  
 CC haematopoietic disorder, platelet disorder or leukaemia; and  
 CC screening of drugs for cancer therapy.  
 CC XX  
 SQ Sequence 1215 AA;

Query Match 93.5%; Score 5860; DB 20; Length 1215;  
 Best Local Similarity 92.9%; Pred. No. 0;  
 Matches 1129; Conservative 39; Mismatches 45; Indels 2; Gaps 2;

Db 361 GNLLEKRRQALLLEQQRKEQERLAQLERAQERKERERQERKROLELEKQRELE 420  
 QY RQREERERKEIERRAARELEERQOLEMERENROELLNQNKEQEGVYVAKARKTLEF 479  
 Db 421 RQREERERKEIERRAARELEERQOLEMERENROELLNQNKEQEOI VVAKARKTLEF 480  
 QY ELEALNDKKHOLEGLQDIRGLATORQEIETSKSRELAETIHLQOQLOESQOMLGR 539  
 Db 481 ELEALNDKKHOLEGLQDIRGLATORQEIETSKSRELAETIHLQOQLOESQOMLGR 540  
 QY LIPEKQIISDQKQOYQONS LHRDSLTITKRALAEKELARQOLEQDVEBETSKQOEI 599  
 Db 541 LIPEKQIISDQKQOYQONS LHRDSLTITKRALAEKELARQOLEQDVEBETSKQOEI 600  
 QY DVENNQLKELEIHSKQOLQORSLEAARLQKQOEKSELELEKQEDACORVQERDKOW 659  
 Db 601 DIFNNQLKELEIHSKQOLQORSLEAARLQKQOEKSELELEKQEDACORVQERDKOW 660  
 QY LEHVOQE-EGPRPKPHBEDRLKEDSVYRKKEAEERAKPEMOKQSLFPHQEPKAT 718  
 Db 661 LEHVOQEDHQRPRKLHEBEKLEKEESVYKKQGEKQOEADKLGRLFHOHQPAPAV 720  
 QY QAPWSTERKGLTISAQSVVYVYRALYPPESRSHDEITIQPDIVMVDESQTEGEGWL 778  
 Db 721 QAPWSTERKGLTISAQSVVYVYRALYPPESRSHDEITIQPDIVMVDESQTEGEGWL 780  
 QY GGELGKGTGWFPANAEKIPENEVPTPAKYTDLTSAAPKALARETAPAPLVTSSEPT 838  
 Db 779 GGELGKGTGWFPANAEKIPENEVPTPAKYTDLTSAAPKALARETAPAPLVTSSEPT 840  
 QY TPNNMADSSSTWPPSSNNKPEETDMDWTAAQPSLTVSAGQIBORSAPFTATAGSSPSP 898  
 Db 839 TPNNMADSSSTWPPSSNNKPEETDMDWTAAQPSLTVSAGQIBORSAPFTATAGSSPSP 900  
 QY VLGQGEKTEGQAQALYWRRAKKNHNLFNKSDVITVLEQDMMWFGBOGQKMPFKSY 958  
 Db 841 VLGQGEKTEGQAQALYWRRAKKNHNLFNKSDVITVLEQDMMWFGBOGQKMPFKSY 960  
 QY VLGQGEKTEGQAQALYWRRAKKNHNLFNKSDVITVLEQDMMWFGBOGQKMPFKSY 960  
 Db 901 VLGQGEKTEGQAQALYWRRAKKNHNLFNKSDVITVLEQDMMWFGBOGQKMPFKSY 960  
 QY VKLISGPVRKSTSIDTGTESPAKRYASPAKPAIPGEEFIAMYYESSEQDGLTFQO 1018  
 Db 959 VKLISGPVRKSTSIDTGTESPAKRYASPAKPAIPGEEFIAMYYESSEQDGLTFQO 1020  
 QY VKLISGPVRKSTSIDTGTESPAKRYASPAKPAIPGEEFIAMYYESSEQDGLTFQO 1020  
 Db 961 VKLISGPVRKSTSIDTGTESPAKRYASPAKPAIPGEEFIAMYYESSEQDGLTFQO 1020  
 QY GGVIVTTRKDDGMWGTGDKSGVFPSPSYVRLKDSSEGTGKSGIAGKPEIKQVASY 1078  
 Db 1019 GGVIVTTRKDDGMWGTGDKSGVFPSPSYVRLKDSSEGTGKSGIAGKPEIKQVASY 1080  
 QY GGVIVTTRKDDGMWGTGDKSGVFPSPSYVRLKDSSEGTGKSGIAGKPEIKQVASY 1080  
 Db 1021 GGVIVTTRKDDGMWGTGDKSGVFPSPSYVRLKDSSEGTGKSGIAGKPEIKQVASY 1080  
 QY AATGPEQTLAPQQLILIRKKNPGGMEGELQARQKROIGWFPANVYKLLSPGTSKITP 1138  
 Db 1079 AATGPEQTLAPQQLILIRKKNPGGMEGELQARQKROIGWFPANVYKLLSPGTSKITP 1140  
 QY TATGPEQTLAPQQLILIRKKNPGGMEGELQARQKROIGWFPANVYKLLSPGTSKITP 1140  
 Db 1081 TATGPEQTLAPQQLILIRKKNPGGMEGELQARQKROIGWFPANVYKLLSPGTSKITP 1140  
 QY TELPKTAQPAVCOVIGYDVTAAQNDDELAFSKOIIINVLNKEPDMWKGKGVSGVGLFP 1198  
 Db 1139 TELPKTAQPAVCOVIGYDVTAAQNDDELAFSKOIIINVLNKEPDMWKGKGVSGVGLFP 1200  
 QY SNVYKLLTDMPSQ 1213  
 Db 1141 SNVYKLLTDMPSQ 1215  
 QY SNVYKLLTDMPSQ 1215  
 Db 1201 SNVYKLLTDMPSQ 1215

RESULT 4  
 AA32155  
 ID AA32155 standard; Protein; 1220 AA.  
 XX  
 XX AA32155;  
 DT 01-FEB-2000 (first entry)  
 XX  
 XX Human SH3D1A protein.  
 DE SH3D1A; human; Down's syndrome; leukaemia; cancer;  
 XX megakaryocytic abnormality; myeloproliferative disorder;  
 XX platelet disorder; neural disorder; thrombocytopenia;  
 KW

KW haematopoietic disorder; cognitive dysfunction; microcephaly;  
 KW lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy.  
 XX Homo sapiens.  
 XX MO9953062-A2.  
 XX 21-OCT-1999.  
 PD 16-APR-1999; 99WC-US08371.  
 XX 16-APR-1999; 98US-0082007.  
 PR 16-APR-1998; 98US-0082007.  
 XX (CEDA-) CEDARS SINAI HEALTH SYSTEM.  
 PA  
 XX  
 PI Korenberg JR, Chen X;  
 XX MPI, 1999-633829/54.  
 DR N-PSDB; AA234571.  
 DR  
 XX  
 PT Nucleic acid from the human SH3D1A gene and its products, useful for  
 PT the diagnosis and treatment of myeloproliferative disorders and  
 PT leukaemia  
 PS Claim 14; Fig 9; 99p; English.  
 XX

CC This sequence represents the protein encoded by the human SH3D1A  
 CC cDNA clone 21 (see AA234571). SH3D1A contributes to the development  
 CC of platelets and the pathogenesis of leukaemias, both in general  
 CC and in particular those involving the megakaryocytic lineage. The  
 CC gene maps to the small candidate region for low platelets on  
 CC chromosome 21. The protein includes SH3 domains and BH domains,  
 CC both associated with protein-protein interactions and the latter  
 CC with maintenance of the cytoskeleton. At least 3 isoforms of  
 CC SH3D1A exist (see AA32154-58). The invention provides methods for  
 CC the diagnosis and treatment of megakaryocytic abnormality,  
 CC myeloproliferative disorder, platelet disorder, acute leukaemia,  
 CC neural disorders, thrombocytopenia, platelet disorder on  
 CC chromosome 21, low platelets in deletion for 21, association of  
 CC gains in chromosome 21 with leukaemias, neural abnormalities,  
 CC dysfunctions and disorders including brain malformations and  
 CC corresponding cognitive dysfunctions, microcephaly, lissencephaly,  
 CC and colpocephaly. Methods are also provided: for suppressing  
 CC cells unable to regulate themselves; screening for a somatic  
 CC alteration in the SH3D1A gene; monitoring the progress and  
 CC adequacy of a treatment; monitoring tumour risk progress or  
 CC megakaryocytic abnormality, myeloproliferative disorder,  
 CC haematopoietic disorder, platelet disorder or leukaemia; and  
 CC screening of drugs for cancer therapy.  
 CC  
 XX

Sequence 1220 AA;

Query Match 93.3%; Score 5847.5; DB 20; Length 1220;  
 Best Local Similarity 92.5%; Pred. No. 0;  
 Matches 1129; Conservative 39; Mismatches 45; Indels 7; Gaps 3;

QY 1 MAQPPPPGSLVMAITVEERAKHDQFLSLKPIAGFINGDARHPFQSGLPQVLAQ 60  
 DB 1 MAQPPPPGSLVMAITVEERAKHDQFLSLKPIAGFINGDARHPFQSGLPQVLAQ 60  
 QY 61 IMAIADNNDGRMDQVEFSIAMKLILKLGQYQPLSTLPPVMKQDVAISSAPAFGIGI 120  
 DB 61 IMAIADNNDGRMDQVEFSIAMKLILKLGQYQPLSTLPPVMKQDVAISSAPAFGIGI 120  
 QY 121 ASMPPLTAAPVPMGSIPIVGMSPITLVSSVPTAAPPPLANGAPPVQPLPAFAHPAATMP 180  
 DB 121 ASMPPLTAAPVPMGSIPIVGMSPITLVSSVPTAAPPPLANGAPPVQPLPAFAHPAATMP 180  
 QY 181 KSSSFSRSGSGSGLNTLQKAOSFDVASAPPAEMVAPOSSRLKYROLFNSHDKTMSGHL 240  
 DB 181 KSSSFSRSGSGSGLNTLQKAOSFDVASAPPAEMVAPOSSRLKYROLFNSHDKTMSGHL 240  
 QY 241 TGPQARTILMOSSLPOAQLASIMWLSIDIDODGKLTAEFLIAMHLIDVAMSGOPLPVLPL 300

DB 241 TGPQARTILMOSSLPOAQLASIMWLSIDIDODGKLTAEFLIAMHLIDVAMSGOPLPVLPL 300  
 QY 301 PEYIPPSFRVRVSGSGSVISSSSVDQRLPEPSSSEDEQOP-EKKLPVTFEDKKRENFER 359  
 DB 301 PEYIPPSFRVRVSGSGSVISSSTVDQRLPEPVLDEEQOLEKKLPVTFEDKKRENFER 360  
 QY 360 GSVLEKRRQALLQQRKEORLAQLERAEOKERERQOEAKROLLELEKOLEKELE 419  
 DB 361 GNLELEKRRQALLQQRKEORLAQLERAEOKERERQOEAKROLLELEKOLEKELE 420  
 QY 420 ROREERERKEIERERAKRELEROROLEMERNRROELLNPNKNEQEDIIVLKAKKTLF 479  
 DB 421 ROREERERKEIERERAKRELEROROLEMERNRROELLNPNKNEQEDIIVLKAKKTLF 480  
 QY 480 ELEALNDKKHOLEKLODIRCLATQROEIESTYKSRHELIAETTHLOOQLOESQMLGR 539  
 DB 481 ELEALNDKKHOLEKLODIRCLATQROEIESTYKSRHELIAETTHLOOQLOESQMLGR 540  
 QY 540 LIPEKQILSDQLKQVOONSJLHRSLLTLKALAEKELARQOLREQULDEVERETRSLQEI 599  
 DB 541 LIPEKQILSDQLKQVOONSJLHRSLLTLKALAEKELARQOLREQULDEVERETRSLQEI 600  
 QY 600 DVFNNQKLELREHSKQOLQKQSLLEAARLKQKEERKSLLEKQEDDAQRVQERDKQW 659  
 DB 601 DVFNNQKLELREHSKQOLQKQSLLEAARLKQKEERKSLLEKQEDDAQRVQERDKQW 660  
 QY 660 LEHQOQ-EQPRPKPHEDELKREDSVRKKEAEERAKPEMODKOSLFFPHOEPKATLAT 718  
 DB 661 LEHQOQ-EQPRPKPHEDELKREDSVRKKEAEERAKPEMODKOSLFFPHOEPKATLAT 720  
 QY 719 QAPMSTTEKGLTISAQSVKVVYRYALYPESSRSHDEITIQPDIVM----VDESQTS 773  
 DB 721 QAPMSTTEKGLTISAQSVKVVYRYALYPESSRSHDEITIQPDIVM----VDESQTS 780  
 QY 774 EPQWIGSLKKGKTGMFPANVAEKIPENEVPPTPAKPVYDLSAPAPKALRETPAPLEVTS 833  
 DB 781 EPQWIGSLKKGKTGMFPANVAEKIPENEVPPTPAKPVYDLSAPAPKALRETPAPLEVTS 840  
 QY 834 SEPSTTPNNMADFSTWSSSNEKEPETDNMDTMAOPLTVPSAGQULROSAFPTATATG 893  
 DB 841 SEPSTTPNNMADFSTWSSSNEKEPETDNMDTMAOPLTVPSAGQULROSAFPTATATG 900  
 QY 894 SSPSPVLQGGKQVGLQALYPMRAKKDNLNFKNSDVITVLEEQDMMWFGEVQGGKW 953  
 DB 901 SSPSPVLQGGKQVGLQALYPMRAKKDNLNFKNSDVITVLEEQDMMWFGEVQGGKW 960  
 QY 954 FPKSYVKLISGPVKKSTSIDTGPTESPASLKRVASPAKPAIPBEERTIANYTSSSQGD 1013  
 DB 961 FPKSYVKLISGPVKKSTSIDTGPTESPASLKRVASPAKPAIPBEERTIANYTSSSQGD 1020  
 QY 1014 LTFQGGDVIIVTKKDGDMWGTGVDKSGSVFSPSNVYRLKDSGSGTAGKTGSLGKKPEIAQ 1073  
 DB 1021 LTFQGGDVIIVTKKDGDMWGTGVDKSGSVFSPSNVYRLKDSGSGTAGKTGSLGKKPEIAQ 1080  
 QY 1074 VIASYATGPEQTLTAPQQLILIRKKNPGWMEGELDARGKKRQIGMFPANVYVLTLSPT 1133  
 DB 1081 VIASYATGPEQTLTAPQQLILIRKKNPGWMEGELDARGKKRQIGMFPANVYVLTLSPT 1140  
 QY 1134 SKITTELPKTAVOPAVCVIGMTDYTAQNDDELAFSGKQIIVNLKNEPDMWKGVEVSGQ 1193  
 DB 1141 SKITTELPKSTALAAVCQVIGMYDYTAQNDDELAFUNKQIIVNLKNEPDMWKGVEVSGQ 1200  
 QY 1194 VGLFSPSNVYKLTMDPDSQ 1213  
 DB 1201 VGLFSPSNVYKLTMDPDSQ 1220  
 RESULT 5  
 ID AA32154 standard; Protein: 1144 AA.  
 XX  
 AC AA32154;



QY 1019 GDVIVTKKDDMMGTGVDKSGVPSNYRLKDSGSGTACKTSGLGKKPEIAQVIYASY 1078  
 Db 1001 -----ETIAQVIASY 1009  
 QY 1079 AATGPOLTLAPGOLLIRKKNPGWMEGELQARGKKROI GMFPANYVYLKSGTSKIRP 1138  
 Db 1010 TATGPOLTLAPGOLLIRKKNPGWMEGELQARGKKROI GMFPANYVYLKSGTSKIRP 1069  
 QY 1139 TELPKTAVGPVAVCOVIGMDYDTAQNDELAFSGKQIINVLNKEDPDMWKGVSQVGLFP 1198  
 Db 1070 TEPKSTALAAVCOVIGMDYDTAQNDELAFNKGQIINVLNKEDPDMWKGVSQVGLFP 1129  
 QY 1199 SNVYKLTMDPSQ 1213  
 Db 1130 SNVYKLTMDPSQ 1144  
 RESULT 6  
 AAU87168  
 ID AAU87168 standard; Protein; 1031 AA.  
 XX  
 AC AAU87168;  
 XX  
 DT 05-JUN-2002 (first entry)  
 XX  
 DE Novel central nervous system protein #78.  
 XX  
 KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;  
 KW hyperproliferative disorder; neoplasm; cardiovascular disorder;  
 KW cardiac arrest; cerebrovascular disorder; ischemia; angiogenesis;  
 KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;  
 KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;  
 KW adenocarcinoma; reproductive system disorder; testicular feminization;  
 KW endocrine disorder; diabetes; cancer; leukemia; neovascularization;  
 KW respiratory disorder; renal disorder; kidney failure; blood disorder;  
 KW myocardial infarction; wound healing; cell proliferation; skin aging;  
 KW food additive; food preservative; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200155318-A2.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US01332.  
 XX  
 PR 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226868.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 21-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234927.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240950.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.

PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249219.  
 PR 17-NOV-2000; 2000US-0249220.  
 PR 17-NOV-2000; 2000US-0249221.  
 PR 17-NOV-2000; 2000US-0249222.  
 PR 17-NOV-2000; 2000US-0249223.  
 PR 17-NOV-2000; 2000US-0249224.  
 PR 17-NOV-2000; 2000US-0249225.  
 PR 17-NOV-2000; 2000US-0249226.  
 PR 17-NOV-2000; 2000US-0249227.  
 PR 17-NOV-2000; 2000US-0249228.  
 PR 17-NOV-2000; 2000US-0249229.  
 PR 17-NOV-2000; 2000US-0249230.  
 PR 17-NOV-2000; 2000US-0250150.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251858.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 XX WPI; 2001-581633/65.  
 DR N-PSDB; ABK43498.  
 XX  
 PT New isolated nucleic acid encoding a protein for diagnosing,  
 PT preventing, treating or ameliorating medical conditions and used as  
 PT food additives or preservatives -  
 XX  
 PS Claim 9; SEQ ID No 686; 837bp; English.  
 XX  
 CC The invention describes an isolated nucleic acid molecule (I) encoding a  
 CC novel central nervous system protein. (I) and polypeptides (II) encoded  
 CC by (I), are used to treat a medical conditions and in diagnosis of a  
 CC pathological condition. Disorders which are diagnosed or treated include  
 CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative  
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders  
 CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischemia,  
 CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and  
 CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses  
 CC e.g. acquired immunodeficiency virus (AIDS) and fungi, ocular disorders  
 CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,  
 CC adenocarcinomas and irritable bowel syndrome, reproductive system  
 CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes  
 CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.  
 CC leukemia, disorders involving neovascularisation e.g. malignancies,  
 CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.  
 CC acute kidney failure and blood related disorders e.g. myocardial  
 CC infarction. The polypeptides can also be used to aid wound healing and  
 CC epithelial cell proliferation, to prevent skin aging due to sunburn, to  
 CC maintain organs before transplantation, for supporting cell culture of  
 CC primary tissues, to regenerate tissues and in chemotaxis. The  
 CC polypeptides can also be used as a food additive or preservative to  
 CC increase or decrease storage capabilities, fat content, lipid, protein,

Query Match 68.2%; Score 4277; DB 22; Length 1031;  
 Best Local Similarity 92.3%; Pred. No. 3,4e-266;  
 Matches 838; Conservative 28; Mismatches 36; Indels 6; Gaps 3;

QY 1 MAQPTPPGSGSIDWATVTEERAKHDOQFLSKPIAGFITDDQARNFFPGGLPQPVLAQ 60  
 DB 8 MAQPTPPGSGSIDWATVTEERAKHDOQFLSKPIAGFITDDQARNFFPGGLPQPVLAQ 67  
 QY 61 IVALADNNNDGRMDQVEFSIAMKLIKLOGYQVLPSTLPPMKOOPVAISSAPAFGIGI 120  
 DB 68 IVALADNNNDGRMDQVEFSIAMKLIKLOGYQVLPSTLPPMKOOPVAISSAPAFGIGI 127  
 QY 121 ASMPPLTAVAPVPMGSIPIVVGMSPPVSSVPPAAVPLIANGAPVPIOPLPFAHPAATWP 180  
 DB 128 ASMPPLTAVAPVPMGSIPIVVGMSPPVSSVPPAAVPLIANGAPVPIOPLPFAHPAATWP 187  
 QY 181 KSSFSRSGPQSGLNTKLOKASPDVASAPPAEAVQSSRLKYROLFNHSDTMSGHL 240  
 DB 188 KSSFSRSGPQSGLNTKLOKASPDVASAPPAEAVQSSRLKYROLFNHSDTMSGHL 247  
 QY 241 T----GPOARTILMOSLPOAQLASIMWLSIDIDGKLTAEFTIAMHLIDVAMSGOLP 296  
 DB 248 TSCCLGPOARTILMOSLPOAQLASIMWLSIDIDGKLTAEFTIAMHLIDVAMSGOLP 307  
 QY 297 PVLPEBYIIPPSFRVRSQSGMSVSSSVDDRLPEBSSDEQOP-EKKLPVTEDEKRE 355  
 DB 308 PVLPEBYIIPPSFRVRSQSGMSVSSSVDDRLPEBSSDEQOP-EKKLPVTEDEKRE 367  
 QY 356 NFERGSVELERKRALLEQORKEORLQLEBRAEOREREROROEAKROLEKOLEKQ 415  
 DB 368 NFERGNLEERKRALLEQORKEORLQLEBRAEOREREROROEAKROLEKOLEKQ 427  
 QY 416 RELERQREERERKEIERRDAKRELEROROLEWENRRQELINQNKQEGTVVLKARK 475  
 DB 428 RELERQREERERKEIERRDAKRELEROROLEWENRRQELINQNKQEGTVVLKARK 487  
 QY 476 TLEFLEALNDKHOLEGLODIRCLATROEISTKSELRLAETTHOQOLOSQ 535  
 DB 488 TLEFLEALNDKHOLEGLODIRCLATROEISTKSELRLAETTHOQOLOSQ 547  
 QY 536 MGRLIPEKQILSDQKOVQNSLHRDSLTLKRALLEAKLAROQLREQDEVERETRSK 595  
 DB 548 MGRLIPEKQILSDQKOVQNSLHRDSLTLKRALLEAKLAROQLREQDEVERETRSK 607  
 QY 596 LQIIVFNQQLKEIRLEHSKOLOKORSLBARLKQXOEKSLLEKXEDAORVOR 655  
 DB 608 LQIIVFNQQLKEIRLEHSKOLOKORSLBARLKQXOEKSLLEKXEDAORVOR 667  
 QY 656 DKWLEHVQOE-EQPRPKPHEEDRLKEDSVRKKEAEERAKPEMODQSRLEFHPQEA 714  
 DB 668 DKWLEHVQOEDEHQRPRKLEHEEKLKREESVKKDGEKEKQOEADQLGRLEFHQHPA 727  
 QY 715 KLTQAPWMTTKKGLTLTSAQSVKVVYVYRALYPRESSHBEITQOPDYMUNDESQGE 774  
 DB 728 KAVQAPWMTAKGGLTLTSAQSVKVVYVYRALYPRESSHBEITQOPDYMUNDESQGE 787  
 QY 775 PGWLGELGKGTGMPFANYAEKIPENEVPTPAKPTDLSAPAPLALRETPAPLPVTS 834  
 DB 788 PGWLGELGKGTGMPFANYAEKIPENEVPTPAKPTDLSAPAPLALRETPAPLPVTS 847  
 QY 835 EPGTTPNNWADSSSWPSSSNEKEPTDWDWTAQPSLTVSAGOLRORSATPATAGS 894  
 DB 848 EPGTTPNNWADSSSWPSSSNEKEPTDWDWTAQPSLTVSAGOLRORSATPATAGS 907  
 QY 895 SPSPVVGQ 902  
 DB 908 SPSPVVGQ 915

RESULT 7  
 AAM43519  
 ID AAM43519 standard; Protein; 1035 AA.

XX AC AAM4519;  
XX 22-OCT-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 197.  
XX  
XX Human; antiarthritic; antirheumatic; antiproliferative; vasotropic;  
XX cerebroprotective; nootropic; neuroprotective; antibacterial; virucide;  
XX fungicide; ophthalmological; cytostatic; immunosuppressive; nootropic;  
XX neuroprotective; anti allergic; hepatocytic; antidiabetic;  
XX antiinflammatory; antiulcer; vulnery; anticonvulsant; antibacterial;  
XX antiparasitic; gardiant; gene therapy; cancer; immune disorder;  
XX cardiovascular disorder; neurological disease; infection; human.  
XX  
XX Homo sapiens.  
XX  
XX WO200155308-A2.  
XX  
XX 02-AUG-2001.  
XX  
XX  
XX 17-JAN-2001; 2001WO-US01309.  
XX  
XX  
XX 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
XX 24-FEB-2000; 2000US-0184664.  
XX 02-MAR-2000; 2000US-0186350.  
XX 16-MAR-2000; 2000US-0189874.  
XX 17-MAR-2000; 2000US-0190076.  
XX 18-APR-2000; 2000US-0198123.  
XX 19-MAY-2000; 2000US-0205515.  
XX 07-JUN-2000; 2000US-0209467.  
XX 28-JUN-2000; 2000US-0214886.  
XX 30-JUN-2000; 2000US-0215135.  
XX 07-JUL-2000; 2000US-0216647.  
XX 07-JUL-2000; 2000US-0216880.  
XX 11-JUL-2000; 2000US-0217487.  
XX 11-JUL-2000; 2000US-0217496.  
XX 14-JUL-2000; 2000US-0218290.  
XX 26-JUL-2000; 2000US-0220963.  
XX 26-JUL-2000; 2000US-0220964.  
XX 14-AUG-2000; 2000US-0224518.  
XX 14-AUG-2000; 2000US-0224519.  
XX 14-AUG-2000; 2000US-0225213.  
XX 14-AUG-2000; 2000US-0225214.  
XX 14-AUG-2000; 2000US-0225266.  
XX 14-AUG-2000; 2000US-0225267.  
XX 14-AUG-2000; 2000US-0225268.  
XX 14-AUG-2000; 2000US-0225270.  
XX 14-AUG-2000; 2000US-0225447.  
XX 14-AUG-2000; 2000US-0225457.  
XX 14-AUG-2000; 2000US-0225758.  
XX 14-AUG-2000; 2000US-0225759.  
XX 18-AUG-2000; 2000US-0226279.  
XX 22-AUG-2000; 2000US-0226681.  
XX 22-AUG-2000; 2000US-0226688.  
XX 22-AUG-2000; 2000US-0227182.  
XX 23-AUG-2000; 2000US-0227009.  
XX 30-AUG-2000; 2000US-0228824.  
XX 01-SEP-2000; 2000US-0229287.  
XX 01-SEP-2000; 2000US-0229343.  
XX 01-SEP-2000; 2000US-0229344.  
XX 01-SEP-2000; 2000US-0229345.  
XX 05-SEP-2000; 2000US-0229509.  
XX 05-SEP-2000; 2000US-0229513.  
XX 06-SEP-2000; 2000US-0230437.  
XX 06-SEP-2000; 2000US-0230438.  
XX 08-SEP-2000; 2000US-0231242.  
XX 08-SEP-2000; 2000US-0231243.  
XX 08-SEP-2000; 2000US-0231244.  
XX 08-SEP-2000; 2000US-0231413.  
XX 08-SEP-2000; 2000US-0231414.  
XX 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233053.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234224.  
PR 25-SEP-2000; 2000US-0234937.  
PR 25-SEP-2000; 2000US-0234938.  
PR 25-SEP-2000; 2000US-0234939.  
PR 26-SEP-2000; 2000US-0234984.  
PR 27-SEP-2000; 2000US-0235836.  
PR 27-SEP-2000; 2000US-0235837.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236328.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249287.  
PR 17-NOV-2000; 2000US-0249289.  
PR 17-NOV-2000; 2000US-0249300.

PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 06-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PI Rosen CA, Barash SC, Ruben SM;  
 DR WPI; 2001-486781/53.  
 DR N-PSDB; AAI63825.  
 XX  
 PT New isolated nucleic acids and polypeptides, useful for diagnosing,  
 PT treating and/or preventing human diseases and disorders -  
 XX  
 XX Claim 11; SEQ ID NO 197; 664pp + Sequence Listing; English.  
 XX  
 CC The invention relates to human polynucleotides (AAI63803-AAI64012) and  
 CC the encoded proteins (AAI634497-AAI63660) useful for preventing, treating  
 CC or ameliorating medical conditions e.g. by protein or gene therapy. The  
 CC genes were isolated from a range of human tissues disclosed in the  
 CC specification. The nucleic acids, proteins, antibodies and (ant) agonists  
 CC are useful in the diagnosis, treatment and prevention of: (a) cancer,  
 CC e.g. breast and ovarian cancer and other cancers of the adrenal gland,  
 CC bone, bone marrow, breast, gastrointestinal tract, liver, lung, or  
 CC urogenital; (b) immune disorders e.g. Addison's disease, allergies,  
 CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,  
 CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 XX Sequence 1035 AA:  
 Query Match 68.2%; Score 4277; DB 22; Length 1035;  
 Best Local Similarity 92.3%; Pred. No. 3,4e-266;  
 Matches 838; Conservative 28; Mismatches 36; Indels 6; Gaps 3;

Db 308 PVLPEYIPSPFRVRSGSGSVSVTSVDQRLPEEPVLEDEQQLEKKLPVTFEDKKRE 367  
 QY 356 NFEFGSVLEKRRQALLLEQKREOERLAQLEPAEOKERKEREDQAKRLLELEKLEKQ 415  
 Db 368 NFERGNLELEKRRQALLLEQKREOERLAQLEPAEOKERKEREDQAKRLLELEKLEKQ 427  
 QY 416 RELEKREERERKEIEREAKRELEROROLEMERNRROELNORNEQSGTVLKARR 475  
 Db 428 RELEKREERERKEIEREAKRELEROROLEMERNRROELNORNEQSGTVLKARR 487  
 QY 476 TLEFELALNDKKQLBGKLDIRCRLATOREIESTTNKSRELRIAETHLQOQLOESQ 535  
 Db 488 TLEFELALNDKKQLBGKLDIRCRLATOREIESTTNKSRELRIAETHLQOQLOESQ 547  
 QY 536 MLEGRILPEKQILNDQLOKVOONSJHRDLSLTAKALEAKELARQHLMDQLEVEKERTSK 595  
 Db 548 MLEGRILPEKQILNDQLOKVOONSJHRDLSLTAKALEAKELARQHLMDQLEVEKERTSK 607  
 QY 596 LOEIDVFNNOQLEKEIREIHSKOOLKORSLEAARLKOEKREKSLLEBKOKEDAQRRVOER 655  
 Db 608 LOEIDVFNNOQLEKEIREIHSKOOLKORSLEAARLKOEKREKSLLEBKOKEDAQRRVOER 667  
 QY 656 DKQWLEHVQOE-EGPRPKPHEEDRLKREDSVKRKEABERAKPMDOKOSRLFFPHQEP 714  
 Db 668 DKQWLEHVQOEDEHQRPKLEHEEKLRKEESVKKKKDEKQKQADKRLRFFHQEP 727  
 QY 715 KLATQAPWSTTEKGPLTISAQESVYVYRALYPFESRSHDEITQGDIVMDESQTGE 774  
 Db 728 KPAVAPWSTTEKGPLTISAQESVYVYRALYPFESRSHDEITQGDIVMDESQTGE 787  
 QY 775 PGMLEGLKGTGWFPANVYAEKIFENEVPTPAKVDTLTSAPAKLALRETPAPLPVTSS 834  
 Db 788 PGMLEGLKGTGWFPANVYAEKIFENEVPTPAKVDTLTSAPAKLALRETPAPLPVTSS 847  
 QY 835 EESTTPNNWAPSSSWPSSSNEKETDNDWMAQPSLTVPSAQOLRORSAFTPATATGS 894  
 Db 848 EESTTPNNWAPSSSWPSTVNEKETDNDWMAQPSLTVPSAQOLRORSAFTPATATGS 907  
 QY 895 SPSPVLCQ 902  
 Db 908 SPSPVLCQ 915  
 RESULT 8  
 ID AAG20575 standard; Protein; 932 AA.  
 XX  
 AC AAG20575;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #20566.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS84762.



XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
PS Claim 20, SEQ ID No 50934, 103pp: English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 932 AA;  
Query Match 49.4%; Score 3097.5; DB 22; Length 932;  
Best Local Similarity 80.0%; Pred. No. 2e-190;  
Matches 603; Conservative 26; Mismatches 26; Indels 99; Gaps 3;

DB 492 FIAMTYESESQGLTFQGGDIVVITKKDGMWGTGTVDKAGVPSNVYRLKDSGSGTA 551  
QY 1060 GTTGSIGKKPEIAOVIASAAATGPQTLAPQQLIRKKNPGWEGELQARKKRIIG 1119  
DB 552 GKTGSIGKKPEIAOVIASAAATGPQTLAPQQLIRKKNP----- 593  
QY 1120 WFPANYVKLLSPGSKITPTELPTKTAQPAVCOVIGWYDPAQNDELAFSKQIINVLN 1179  
DB 594 -----VCQVIGWYDPAQNDELAFNKGQIINVLN 623  
QY 1180 KEDPDMWKGEGVSGVGLPSPNVYKLTMDPSQ 1213  
DB 624 KEDPDMWKGEGVSGVGLPSPNVYKLTMDPSQ 657  
RESULT 9  
AA017881  
ID AA017881 standard; Protein, 1681 AA.  
XX  
AC AA017881;  
XX  
DT 22-AUG-2002 (first entry)  
XX  
DE Allergic disease examination method related human protein.  
XX  
KW Allergic disease; allergy; antiallergic; intersectin 2; eosinophil;  
XX atopic dermatitis; human.  
XX Homo sapiens.  
XX OS  
XX PN WO200233122-A1.  
XX  
PD 25-APR-2002.  
XX  
PF 11-OCT-2001; 2001WO-JP08937.  
XX  
PR 13-OCT-2000; 2000JP-0314093.  
XX  
PA (GENO-) GENOX RES INC.  
PA (NIGE-) JAPAN GEN NAT CHILDREN'S HOSPITAL.  
PA (EISA) EISAI CO LTD.  
XX  
PI Sugita Y, Hashida R, Ogawa K, Obayashi M, Nagasu T, Saito H;  
XX Takahashi E;  
XX  
DR MPI: 2002-372313/40.  
XX  
DR N-PSDB: AAL47247.  
XX  
PT Method for examining allergic diseases by differential display of  
PT intersectin 2 gene showing different expression particularly  
PT significant increase in eosinophils in patients  
XX  
PS Disclosure: Page 78-86; 90pp; Japanese.  
XX  
XX The present invention relates to a method for examining allergic diseases  
XX CC with intersectin 2 gene or a gene with equivalent function of intersectin  
XX CC 2 as an indicator gene, which comprises determining the expression level  
XX CC of the gene in the eosinophils in a patient, and comparing the expression  
XX CC level with that in the eosinophils of a healthy individual. The method is  
XX CC also applicable in screening candidate compounds for remedies. The  
XX CC present sequence is a human protein described in the exemplification of  
XX CC the invention.  
SQ Sequence 1681 AA;  
Query Match 49.3%; Score 3091.5; DB 23; Length 1681;  
Best Local Similarity 52.1%; Pred. No. 1.1e-189;  
Matches 656; Conservative 187; Mismatches 274; Indels 143; Gaps 29;

```

Db      13 MAQPTAMNGCPNNMATTSEETKRDROFNLRKSGGYITGDQARNFELQSLPAPVLA 72
Qy      61 IVALADNNNGRMDOVEFFSIAMKLIKLOGYOLPSTLPVPMKQPV--AISSAPAFGIG 118
Db      73 IMAISDLNKGKMQQOFESIAMKLIKLOGOQOLPVLPIMQPPMFSLISA-RFGMG 131
Qy      119 GIASMPPLTAVAPVPMG-----SIPVGMSPVLVSVPPAAPPVLANGAPV 165
Db      132 ---SMPNLISIPOLPAPAPITSLSSATSGTNLPMLMPTLPVSPVSTSLP---NGTASL 185
Qy      166 IQPLPAFAHAPATWPKSSSF--RSGGSQLNTKLOAKS--PDVASAP----- 211
Db      186 IQPLP-IPYSSSTLPKSSSYSLMMGFG--GASIQAGSLIDGSSSSSTSLASGNS 241
Qy      212 ---AAEWAPOSSRLKYLXROLFNSHDKTMSGHLTGPOARTLMQSSLPQALASIMNLS 267
Db      242 PKGTSEMAVAPQPTRLKYRQKFNLTLDKSMGYSLSGQARNALQSNLSQOLATIMTLAD 301
Qy      268 IDQDGKLTAEFFILAMHLDVAMSGOLPVLPPREYIPBSFRFRSGSGMSVSSSDVDQ 327
Db      302 VDGGLKAEFFILAMHLDMAKAGOLPLPLPELVPPSFR-----GKQOI--DSING 353
Qy      328 RLPEPSEDEQOEKLPVTFEDKKRNPERGSVELEKROALLEQORKEORLAQLER 387
Db      354 TLPSYQMG-EEFQKLPVTFEDKKRANYPGRMELFKRRQALMEQOQREARKQKEX 412
Qy      388 AEOERKERQEOEAKOLELEKOLEKORERQEEERKEIERBAKRELEKOROLE 447
Db      413 EEMERKOREIOEOMKOLELEKOLEKORERQEEERKDIERRBAQOELEKORHLE 472
Qy      448 WERRRRELLNORMKOEQGVVLKARRKLEFELEALNDKKHOLEGLLOIRCLARQ 507
Db      473 WERRRRELLNORMKOEQBEIYRLNSKKNLHLEBALNGHQOISGLQDVRLLKQIQKT 532
Qy      508 EIESTNKSRELIATHTLOOLOESQOMGLRLPEKOLISDLQKOY-ONSJHRDSLLT 566
Db      533 ELEVLDKQCDLEIMEIKLOQOELOEQNKLIYLVPEKOLLNERIKNNQFSTPDSGSL 592
Qy      567 LKRALLEKELARQOLREOLDEVERETRSKLOEIVFNNOLEKLEHLSKQOLKORSL 626
Db      593 HKKSLKEKELC-ORLKEOLDALKEKETASKLSEMSFNNOLEKETETNTQOLALEQYKI 651
Qy      627 ARLKOEKERSLELEK-----EDAQRVQERDKOMLEHVOEOPRPKPHEDLTKR 682
Db      652 KRDKLKEIERKLELMQKKLEDBAKAKQKGNLKENIRKEE-----EKQKRLQE 705
Qy      683 EDSVRKKEAEERAKREMOKOSRLFNHQBPAKLATQAPWSTTEKPLTISAQSVK-- 739
Db      706 EKTQEKIOEERKABEKORKD-----TLKAEKKRETA 740
Qy      740 --VYYRRLYPPESRSHETITQBGDIYVNDSESOTGERGWLGBLKGKTGFNPANVAK 797
Db      741 SVLVNYRALYPPFARNHDEMSFGSDIIOVEKTVGEEGMWYGSFGQNFQGFNPANYEKM 800
Qy      798 P--ENEVTPAKPVTDLTASAPAKLALRETPAPLPVTSSEBSITPNWADFFSSTWPS 855
Db      801 PSSENE-----KAVSPKAL-----LPPTJSLATS-----TSEPLESSN 835
Qy      856 EKPEPTNDWTAAOQSLTVPAGOLRQSAFTPATGSSPYLVGQOEKVEGICQAOALY 915
Db      836 QPASVTDQON--VSFSLNLTVNTSWQ--KKSAPTRTVSPG--SVSPFHOGQOVVENIKAOALC 891
Qy      916 FWRKKDNLHLPNKSVDVTLVEQDDMMWFGVQOKGFPKSYKLTLSGPRKSTSDTG 975
Db      892 SWTAKDNLHLPNKSVDVTLVEQDDMMWFGVHGGKGFPSYKLTIPGSVK----- 944
Qy      976 PTESPASLKRVAS--PAKPAIPGEEFLAMTYESSSEOGDLPFOQGDVIVTKKGDWMT 1033
Db      945 -REPEALYAIVNKKPTSAASVGEVIALYPSVSEBGLDTTEGEBEILVTKDGEWMT 1003
Qy      1034 GTVGDKSVFVFSNRYRLDSESGAGTSGSLGKKPEIAQVITASYATATGPQULTLAGOL 1093
Db      1004 GSGIDRSGIFPSNYVKPKDOESFGSASKSGANKKPEIAQVITASYATVYVSGSEQSLAGOL 1063

```

```

Qy      1094 IIRKKNPGWMEGBLDRKKROIGWFPANYVKLSPPGTSKITPTLEPTANOVANCOV 1153
Db      1064 ILILEKNTSGWMOGELQARKKRQKGFPAASHVXLPLPSSRBAPARP-----VCQV 1116
Qy      1154 IGMVYTAQNDELAFSGQIIVNLNKEDPDMWKEVSGQVGLFPSSNYVKLTMDPSQ 1213
Db      1117 IAMDYAAANNDELSFGKQOLINWANKDDPDMQGEINGVTGLFPSSNYVKLTMDPSQ 1176

RESULT 10
AAV32157
ID AAV32157 standard; Protein: 648 AA.
XX
AC AAV32157;
XX
DT 01-FEB-2000 (first entry)
XX
DE Human SH3D1A protein.
KW SH3D1A; human; Down's syndrome; leukaemia; cancer;
KW megakaryocytic abnormality; myeloproliferative disorder;
KW platelet disorder; neural disorder; thrombocytopenia;
KW haematopoietic disorder; cognitive dysfunction; microcephaly;
KW lissencephaly; colpocephaly; holoprosencephaly; diagnosis; therapy.
XX
OS Homo sapiens.
XX
PN W09953062-A2.
XX
PD 21-OCT-1999.
XX
PF 16-APR-1999; 99WO-US08371.
XX
PR 16-APR-1998; 98US-0082007.
XX
PA (CEDA-) CEDARS SINAI HEALTH SYSTEM.
XX
PI Korenberg JR, Chen X;
XX
DR WPI, 1999-633829/54.
XX
N-PSDB; AAZ34573.
XX
PT Nucleic acid from the human SH3D1A gene and its products, useful for
PT the diagnosis and treatment of myeloproliferative disorders and
PT leukaemia.
XX
PS Claim 14; Fig 13; 99pp; English.
XX
XX
XX
This sequence represents the protein encoded by the human SH3D1A
CDNA clone 5 (see AAZ34573). SH3D1A contributes to the development
of platelets and the pathogenesis of leukaemias, both in general
and in particular those involving the megakaryocytic lineage. The
gene maps to the small candidate region for low platelets on
chromosome 21. The protein includes SH3 domains and EH domains,
both associated with protein-protein interactions and the latter
with maintenance of the cytoskeleton. At least 3 isoforms of
SH3D1A exist (see AAV32154-58). The invention provides methods for
the diagnosis and treatment of megakaryocytic abnormality,
myeloproliferative disorder, platelet disorder, acute leukaemia,
neural disorders, thrombocytopenia, platelet disorder on
chromosome 21, low platelets in deletion for 21, association of
gains in chromosome 21 with leukaemias, neural abnormalities,
dysfunctions and disorders including brain malformations and
corresponding cognitive dysfunctions, microcephaly, lissencephaly,
and colpocephaly. Methods are also provided: for suppressing
cells unable to regulate themselves; screening for a somatic
alteration in the SH3D1A gene; monitoring the progress and
adequacy of a treatment; monitoring tumour risk progress or
megakaryocytic abnormality; myeloproliferative disorder,
haematopoietic disorder, platelet disorder or leukaemia; and
screening of drugs for cancer therapy.
XX

```

SQ Sequence 648 AA;  
 Query Match 49.0%; Score 3070.5; DB 20; Length 648;  
 Best Local Similarity 94.1%; Pred. No. 6,6e-189;  
 Matches 609; Conservative 20; Mismatches 17; Indels 1; Gaps 1;

QY 1 MAQPTPFGSLDVAIVTEERAKHDOQFLSLKPIAGFITGGDARNFPQSGLPQVLAQ 60  
 DB 1 MAQPTPFGSLDVAIVTEERAKHDOQFLSLKPIAGFITGGDARNFPQSGLPQVLAQ 60  
 QY 61 IVALADNNNDGRMDQVDFSIAMKLIKULOGYLPSTLPVVMKQOQVAISSAPAFIGIG 120  
 DB 61 IVALADNNNDGRMDQVDFSIAMKLIKULOGYLPSTLPVVMKQOQVAISSAPAFIGIG 120  
 QY 121 ASMPPLTAAPVPMGSIPIVVGMSPLVSSVPPAAPPPLANGAPVPIQPLPAFAHPAATWP 180  
 DB 121 ASMPPLTAAPVPMGSIPIVVGMSPLVSSVPPAAPPPLANGAPVPIQPLPAFAHPAATWP 180  
 QY 181 KSSSFSGSGSGLNTKLOAOSFDVASAPPAEMAVPQSSRLKRYQLNNSHDKTMGHL 240  
 DB 181 KSSSFSGSGSGLNTKLOAOSFDVASAPPAEMAVPQSSRLKRYQLNNSHDKTMGHL 240  
 QY 241 TGQARTILMOSSLPOAQLASINWLSIDIDDDGKLTAEFTLAMHLIDVAMSGQPLPVL 300  
 DB 241 TGQARTILMOSSLPOAQLASINWLSIDIDDDGKLTAEFTLAMHLIDVAMSGQPLPVL 300  
 QY 301 PEYIPSFRRVRSGSGMSVSISSSVDOQLPEEPSSEDEQOP-EKKLPVTFEDKKRENFER 359  
 DB 301 PEYIPSFRRVRSGSGMSVSISSSVDOQLPEEPSSEDEQOP-EKKLPVTFEDKKRENFER 359  
 QY 360 GSVELEKRRALLLEQQRKEQERLAQLERABQEKREKEROEQEAKKROLLEKOLEKRELE 419  
 DB 360 GSVELEKRRALLLEQQRKEQERLAQLERABQEKREKEROEQEAKKROLLEKOLEKRELE 419  
 QY 420 ROBEERERKEITEREAKRELEROROLEWERNRROELNORNEQSGTVLAKRRKTLTF 479  
 DB 420 ROBEERERKEITEREAKRELEROROLEWERNRROELNORNEQSGTVLAKRRKTLTF 479  
 QY 480 ELBALNDKKGLEGLQDRIICRLATQROEIESTNKSRELEIAITHLOOOLQESQOMLGR 539  
 DB 480 ELBALNDKKGLEGLQDRIICRLATQROEIESTNKSRELEIAITHLOOOLQESQOMLGR 539  
 QY 540 LIREKQILSOLKOVQONSILHRSLSLTLLKALAKELARQOLREQLDEVERETRSLQEI 599  
 DB 540 LIREKQILSOLKOVQONSILHRSLSLTLLKALAKELARQOLREQLDEVERETRSLQEI 599  
 QY 600 DVNNOLKEIREIHSKQOLOKORSLEAARLKQEKOEKRSLELEKOE 646  
 DB 601 DVNNOLKEIREIHSKQOLOKORSLEAARLKQEKOEKRSLELEKOE 646  
 RESULT 11  
 AAM79199  
 ID AAM79199 standard; Protein; 1697 AA.  
 AC AAM79199;  
 DT 06-NOV-2001 (first entry)  
 DE Human protein SEQ ID NO 1861.  
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KM tissue growth factor; immunomodulatory; cancer; leukaemia;  
 XX nervous system disorder; arthritis; inflammation.  
 OS Homo sapiens.  
 PN MO200157190-A2.  
 PD 09-AUG-2001.  
 PF 05-FEB-2001; 2001WO-US04098.

XX 03-FEB-2000; 2000US-0496914.  
 PR 27-APR-2000; 2000US-0560875.  
 PR 20-JUN-2000; 2000US-0598075.  
 PR 19-JUL-2000; 2000US-0620325.  
 PR 01-SEP-2000; 2000US-0654936.  
 PR 15-SEP-2000; 2000US-0663561.  
 PR 20-OCT-2000; 2000US-0693325.  
 PR 30-NOV-2000; 2000US-0728422.  
 PA (HYSE-) HYSEQ INC.  
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
 PI Xue AJ, Yang Y, Wejntman T, Goodrich R;  
 DR WPI: 2001-476283/51.  
 DR N-PSDB; AAK52332.  
 PT Nucleic acids encoding polypeptides with cytokine-like activities,  
 useful in diagnosis and gene therapy -  
 PS Claim 20; Page 4246-4249; 6221pp; English.  
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAM78123-AAM80302) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoietic regulatory  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
 CC (AAM80020) are omitted as the relevant pages from the sequence listing  
 CC were missing at the time of publication.  
 SQ Sequence 1697 AA;  
 Query Match 48.9%; Score 3068; DB 22; Length 1697;  
 Best Local Similarity 51.0%; Pred. No. 3,6e-188;  
 Matches 656; Conservative 187; Mismatches 274; Indels 170; Gaps 30;

QY 1 MAQPTPFGSLDVAIVTEERAKHDOQFLSLKPIAGFITGGDARNFPQSGLPQVLAQ 60  
 DB 2 MAQPTPFGSLDVAIVTEERAKHDOQFLSLKPIAGFITGGDARNFPQSGLPQVLAQ 61  
 QY 61 IVALADNNNDGRMDQVDFSIAMKLIKULOGYLPSTLPVVMKQOQVAISSAPAFIGIG 118  
 DB 62 IVALADNNNDGRMDQVDFSIAMKLIKULOGYLPSTLPVVMKQOQVAISSAPAFIGIG 120  
 QY 119 GIASMPPLTAAPVPMGSIPIVVGMSPLVSSVPPAAPPPLANGAPVPIQPLPAFAHPAATWP 165  
 DB 121 --SMPNLSIPQPLPAAPATISLSASATSGTNLPLPAMPPTPLVPSVSTSLP--NGTASL 174  
 QY 166 IQPLPAFAHPAATWPKSSSFS--RSGPQSLNTKLOAOS--FDVASAP----- 211  
 DB 175 IQPLP-IPYSSSTLPHGSSSYLMMGGRF--GASIQRAQSILIDGSSSSTSTASLSGNS 230  
 QY 212 ----AAEWAPOSSRLKRYQLNNSHDKTMGHLTGPOARTILMOSSLPOAQLASINWLSID 267  
 DB 231 PKTGSERAAVQPIRLKRYQLNNSHDKTMGHLTGPOARTILMOSSLPOAQLASINWLSID 290  
 QY 268 IDQDGLTAEFTLAMHLIDVAMSGQPLPVLPEYIPPEPSRRVRSGSGMSVSISSSVDO 327  
 DB 291 VDQDGLTAEFTLAMHLIDVAMSGQPLPVLPEYIPPEPSRRVRSGSGMSVSISSSVDO 342  
 QY 328 RLPEEPSSEDEQOP-EKKLPVTFEDKKRENFERGVELEKRRALLLEQQRKEQERLAQLER 387  
 DB 343 TLPSYQKMG-EERPOKLLPVTFEDKKRENFERGVELEKRRALLLEQQRKEQERLAQLER 401

```

QY 388 AEOEREREOEPAROLELEKOLEKORLEFQREBERKEIEREDBAKREIEROROLE 447
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 402 EEWERKORELEOEBWKOLELEKLEKORLEKORLEBEREKDERERBAKORLEORLE 461
QY 448 WEENROELLNONKKEGTVLKKARKTLEFLEALNDKHOLEGLOIRCLLATORO 507
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 462 WERROELLNONKKEGTVLKKARKTLEFLEALNDKHOLEGLOIRCLLATORO 521
QY 508 EIESTNSKREIRIAETITLLOOLOESQOMLGRLLPEKOLISDOLKOVQ-ONSIRHSDLLT 566
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 522 ELEVLDRKOCOLEIMEIKOLELOEYONKLIYIVPEKOLINERIKMOMFNTDSDGVSLL 581
QY 567 LKALAEKELARQOLREOLDEVERETSKEIETDVFNNO----- 605
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 582 HKKSLEKEELC-ORLKQOLDALEKETASKLESDMSFNNOQKCNMDSDVLCILLISLSCI 640
QY 606 -----LKEIREIHSKQOLOKORSLAARLUKOEKORSEKLEKOK-----EDAORVQER 655
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 641 NNLFLLKEIREITNTQOLALBOUYKIKRKLEIERKRELELMOKKLEDEARAKKAKQK 700
QY 656 DKOWLEHVQOEEOPRPRKPHEDRLKREDSVRKKEAEERAKPEMOKOSRLLFPHOEPK 715
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 701 ENLMKENLRKEE-----EKOKRLOEKTQOEKIQEERKAEKOROKD----- 744
QY 716 LATQAPMSTTEKGPLTISAQESVK-----VVYYPALYPPRSRSHDETTIQPGIWWDES 770
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 745 -----TLKAEKKREKTASVLVNTYRATYFPAKRNDEWSGDIIOVDER 789
QY 771 QTEGPGWLGELKGTGMFPANVAEKIP--ENEVTPRAKPVTLTAPAKALRETAP 828
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 790 TVGPGWLYSGFQGNFQMPENYVKMPSSENE-----KAVSPKAL----- 831
QY 829 LPVTSSPESTPNNWADPFSSSTWPSSENEKPEPTDNMTMAOPSLTVPSAQOLQRSAFTP 888
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 832 LPPVTSLSATS-----TSSEPLSSNQPPASVYTDYQ--VSFNSLTVNTSMQ--KKSAPTR 881
QY 889 ATATGSSPSPVLQGEKEVEGLOALYPMWAKKDNHLPNKSPDITVLEQODMMWEGEVQ 948
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 882 TVSGG-SVSPFHGGGVVENKQALCSWTRAKKDNHLPNFKHDITVLEQODMMWEGEVH 940
QY 949 GQKMPFKSYKLISGVRKSTSIDTPTSPASLKRVA--PAKPAIFGEPIAMTYT 1006
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 941 GGRGMPFKSYKIRPGESEVK-----REEPEALYAAVKKPPTSAASYGEEETIALY 992
QY 1007 ESSQGLTFPOGGDVIVYTKKDGDMWTGYGDKSGVPSNVYRLKDESGGTGKGTSLG 1066
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 993 SSVPGDLTFEGEELVTQKDGEMWTGSDGRGIFPSNVYKQDESGSASKSASN 1052
QY 1067 KKPETIAOVIASYATGPEQLTLAPGOLILRKKNPGMWEGELQARGKROIGMFPANVY 1126
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1053 KKPETIAOVTSAYVASSGSEQLSLAPGOLILKKNPSGWMGELQARGKROKGMFPASHV 1112
QY 1127 KLSFGSKITPTLELPTAVOPAVCOYIGMDYTAQNDDELAFSGQIIVNLKEDPDW 1186
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1113 KLDPSSERRATPAHP-----VCQIAYAMYDANNEDLSFSKQILINVMKDDPDW 1165
QY 1187 KGEVSGQVGLPSPNVYKLTMDPSQO 1213
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1166 QGEITNGTGLPSPNVYKLTMDPSQO 1192

```

```

RESULT 12
ID AAY57445
AAV57445 standard; Protein; 1197 AA.
AC AAY57445;
XX 28-FEB-2000 (first entry)
DT
DE Mouse Ese2 protein sequence.
XX
KM Mouse; murine; Ese1; Ese2; endocytosis; vesicular trafficking;
KW regulation; actin cytoskeleton; detection; cancer; infection;

```

```

KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
KM antiproliferative; antiviral.
XX
OS Mus sp.
XX M09955728-A2.
XX
PD 04-NOV-1999.
XX
PF 27-APR-1999; 99MO-CA00375.
XX
PR 27-APR-1998; 98CA-2230201.
XX
PR 05-FEB-1999; 99US-0118739.
XX
PA (HSCR-) HSC RES & DEV LP.
XX
PI Egan SE, Mang W, Sengar A;
DR WPI; 2000-052802/04.
DR N-PSDB; AAZ39010, AAZ39011.
XX
XX New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
PT of endocytosis, used e.g. for treating cancer or preventing viral
XX infection.
XX
PS Claim 33; Page 48; 99pp; English.
XX
XX The present sequence represents mouse Ese2. The present invention
CC specifically describes mammalian Ese1 and 2 proteins (I) and their splice
CC variants (Ese = EH-domain and SH3-domain regulator of endocytosis). (I)
CC are involved in regulation of clathrin-mediated endocytosis (as a complex
CC with Esg1 protein), vesicular trafficking and actin cytoskeleton.
CC Generally (I) (or its (ant)agonists, mimetics, fragments and inactive
CC mutants), (I)-specific antibodies (Ab); sequences antisense to the (I)
CC polynucleotide; agents that downregulate expression of Ese genes or
CC antagonists of an Ese binding partner are used to treat diseases
CC associated with undesirable endocytosis and resulting changes in cellular
CC function. Particularly overexpression of Ese1 is used to block
CC clathrin-mediated endocytosis in vivo or in cell cultures, while
CC administration of (I) is used to promote endocytosis of selected cells.
CC (ant)agonists of (I) or Ab are used to suppress abnormal proliferation of
CC cells that can be stimulated to proliferate by a growth factor receptor,
CC and similar compounds (also inactive Ese mutants) can be used to prevent
CC viral infection. Endocytosis may also be regulated, in vivo or in cell
CC cultures, by forming an Ese-Esg1 complex, then binding dynamin to the
CC complex. Generally conditions that can be treated include cancer;
CC abnormal cell division or migration; viral infection; or abnormal
CC receptor signalling, tissue development or synaptic transmission.
XX
SQ Sequence 1197 AA;
XX
Query Match 48.6%; Score 3049.5; DB 21; Length 1197;
Best Local Similarity 52.4%; Pred. No. 3.5e-187;
Matches 658; Conservative 177; Mismatches 275; Indels 145; Gaps 32;
QY 1 MAQPTPFGSGIDVWATVVERARKDQFLSLKPIAGFITDQARNFPFGSLPQVLAQ 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1 MAQPTPAMNMGGMWMAITSEBRTKHQDFNMLKPSGGVITTDQARTFLOGLAPVLA 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 IVALDMMNDGMDQERSIAMKLIKXLOGYOLPSTLPYMKQOPV--ATSSAPARGIG 118
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 61 IVALSDLNKDKMDQERSIAMKLIKXLOGYOLPSTLPYMKQOPV--ATSSAPARGIG 119
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 119 GIASMP-----PLTAVAPV--PMG-----SIPVGMSPPLVSSVPPAAPPANAP 164
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 120 ---SMPNLSIHQPLRPVAPVPIATPLSSATSGTSLPMLMPALVSVSSLP--NGTAS 173
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 165 VIQPLPAPAPAPATWPKSSSRS--RSGGSQLNTKLQQAQ--FDVYASAP----- 211
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 174 LIQPL-SIPYSSSTLPHASYSYLMWGGFG---GASIQRAQSLIDLGSSSSTASISGN 229
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 212 -----AAEWAVQSSRLYRQLFNSHDKTWSGHLTGPAFTILMOSSLPQQLASIMNLS 266
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db 230 SPPTGSENAVPOPSRLKTRQKFNSLDKMGVLSGFQARMLLSGTLATITWLA 289
Qy 267 DIDODKLTAEFEIILAMHLIDVAMSGQPLPVLPPEYIPPSFRVYSGSGMSVSSSYD 326
Db 290 DIDGDDQLAAEFILAMHLIDMAKAGQPLPLTIPPELVPSFR-----GKGV---DSYN 341
Qy 327 QRLPEPSSSEDEQPEKPLVTFEDKKRNFEBGVSLEKRRQALIEQQRKEORLAQLE 386
Db 342 GTLPYQKQV-EEEPQKPLVTFEDKKRKNYERGMNELKRRQVLMEEQQQRAEKRAQKE 400
Qy 387 RAEOEKREEROEAKROLELEKOLEKORLEKOREEERREKEIEREAKELERQRL 446
Db 401 KEWEKRORELOEWEKKOLEKLEKREKOREEERREKEIEREAKOLEERQRL 460
Qy 447 EEMERNROELLNORNEOEGLTVLAKARKTLEFELEALNDKHQLEGGKQDTRCLATOR 506
Db 461 EEMERLARQELLSQKTEQEDIVRLSRKSLHLELEAVNGKHQISGRQDVQIRKQTK 520
Qy 507 QEIESTNKRRELIATITHLQOOLQOSQOMGLRLPEKQILSDQLKQVQNSLHSDSLT 566
Db 521 TELEVLDKOCDELIMEIKOLOQELKEYQNKLIYLVPEKOLLNERIKNMQLSTNTPSGISL 580
Qy 567 LKRALEAKELARQOLEQLEDEVRETRSKLOEIDVNNQLEKREIHSKQOQOKORSLEA 626
Db 581 LHKSSSEKELCQRLKEQDLEKETAASKLSESDSFNNQLEKREIHSKQOQOKORSLEA 640
Qy 627 ARLKQOEKRSLE-LEKOK---EDARRVOERDKQMLEHVOEOEOPRPRKHEEDRLK 682
Db 641 KNDKLEIKERKLEJOYOKKLEDEAKAKKOGKENMRSIRKEE-----EKQKRLQ 694
Qy 683 EDSVRKKEAEERAKPEMOKOSRLFHPHOEPAKLATQAWSTTEKGPLTISAQESKVVY 742
Db 695 EKSQDTQOEER---KAEAKOSE-----TASA-----LVN 721
Qy 743 YRLAYFEERSHDEITIQGDIYMWDESOTGERGWLGLSKGTGFMFNVAKETIPENV 802
Db 722 YRLAYFEERNHDEMSFGSDIIOVEKTVGEGWLYGSFQCGFPGFPCNYVEKYLSE- 780
Qy 803 PTPAKPVTLTSPAPKALRETAPLPTSSSPPTPNMMAFSSSTWSSSEKRETON 862
Db 781 -----KALSPKALLPPTVLSLSTIS- SQPPASVTDYAHV--SFSNLTVNT-- 824
Qy 863 WDTMAAPSLTVPSAGLQBSAFTPATATGSSPSVLQGEKEVEGLQAOALYPRAKKD 922
Db 825 --TW-----QOKSAFTRTVSG-SVSPIHQGGQVENVLKAQALCSMTAKKE 867
Qy 923 NNLNFKSDVITYLEQDDMMWFGVEVGQKGMFPKSYVKLISG-PVAKS--TSIDGPTES 979
Db 868 NNLNFKSHDVIITYLEQDDMMWFGVEVGQKGMFPKSYVKLIPGNEVORGEPEALYAAVTCK 927
Qy 980 PALKKRVASPAKPAIP-GEEFIAMTYTESSEGGDLTPQGGDIVVTKKDGMMTGTVD 1038
Db 928 PIS---TAIPVISTATVPVEDIYALYSYSVEFGDLTFEGEILVLTQDGEWMTGSIE 984
Qy 1039 KSGVPSNVYRLKDSBGSCTAGTGLGKKPELIAQVIAVYATGPBQTLTABQULIRK 1098
Db 985 RTGIFPSNVYVRPKQENFPNGASKSGASNNKPELIAQVTSAYASGTRQLSLAQULILIK 1044
Qy 1099 KNPNGWBEGLQARKKROIGMPFANYVVLISFGTSKITPTIELPKTAVPANCQYIAMD 1158
Db 1045 KNTSGWMOGELQARKKROIGMPFASHVXLKLGSSSERTPT-----PHAQCOVIAMD 1097
Qy 1159 YTAQNDDELAFSGQIILNKLKEDPPMKWGEVSGQVGLPFSNVYVLTMDPSQ 1213
Db 1098 YMANNEDELNFSGQILNWKNDPDMWOGETNGLTGLPFSNVYVLTMDPSQ 1152

```

```

RESULT 13
AA57450
ID AA57450 standard; protein; 1658 AA.
AC
XX AA57450;
XX

```

```

DT 28-FEB-2000 (first entry)
XX
XX Mouse Ese2L protein sequence.
DE
XX Mouse; murine; Ese1; Ese2; endocytosis; vesicular trafficking;
KW regulation; actin cytoskeleton; detection; cancer; infection;
KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
KW antiproliferative; antiviral.
XX
OS Mus sp.
XX
XX M09955728-A2.
XX
XX 04-NOV-1999.
XX
XX 27-APR-1999; 99MO-CA00375.
XX
XX 27-APR-1998; 98CA-2230201.
XX
XX 05-FEB-1999; 99US-0118739.
XX
XX (HSCR-) HSC RES & DEV LP.
XX
XX Egan SE, Mang W, Sengar A;
XX
XX WPI; 2000-052802/04.
XX
XX N-PSDB; AA239026, AA239027.
XX
XX New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
XX of endocytosis, used e.g. for treating cancer or preventing viral
XX infection.
XX
XX Claim 33; Page 69-70; 99pp; English.
XX
XX
XX The present invention specifically describes mammalian Ese1 and 2
XX proteins (I) and their splice variants (Ese = EH-domain and SH3-domain
XX regulator of endocytosis). (I) are involved in regulation of clathrin-
XX mediated endocytosis (as a complex with Eps15 protein), vesicular
XX trafficking and actin cytoskeleton. Generally (I) (or its (ant)agonists,
XX mimetics, fragments and inactive mutants); (I)-specific antibodies (Ab);
XX sequences antisense to the (I) polynucleotide; agents that downregulate
XX expression of Ese genes or antagonists of an Ese binding partner are
XX used to treat diseases associated with undesirable endocytosis and
XX resulting changes in cellular function. Particularly overexpression of
XX Ese1 is used to block clathrin-mediated endocytosis in vivo or in cell
XX cultures, while administration of (I) is used to promote endocytosis of
XX selected cells. (Ant)agonists of (I) or Ab are used to suppress abnormal
XX proliferation of cells that can be stimulated to proliferate by a growth
XX factor receptor; and similar compounds (also inactive Ese mutants) can be
XX used to prevent viral infection. Endocytosis may also be regulated, in
XX vivo or in cell cultures, by forming an Ese-Eps15 complex, then binding
XX dynamin to the complex. Generally conditions that can be treated include
XX cancer; abnormal cell division or migration; viral infection; or abnormal
XX receptor signalling, tissue development or synaptic transmission. The
XX present sequence represents mouse Ese2L protein sequence.
XX
XX Sequence 1658 AA;

```

```

Query Match 48.6%; Score 3049.5; DB 21; Length 1658;
Best Local Similarity 52.4%; Pred. No. 5,5e-187;
Matches 658; Conservative 177; Mismatches 275; Indels 145; Gaps 32;

```

```

Qy 1 MAOPFPFGSLDVAITVEERAKHQDFLSKPIAGFINGDARNFPOSGLPQPLAQ 60
Db 1 MAOPFPFMNGSPNNMALTSEERTKHDQFNLKPSGGYITIGDQARTFLOSGLPAPVLAE 60
Qy 61 IVALADNNDRMDQVEFSIAMKLILKLGQYQPLSTLPVPMKQQPV--AISSAPAFGIG 118
Db 61 IVALSDLNKQKMDQVEFSIAMKLILKLGQYQPLSTLPVPMKQQPV--AISSAPAFGIG 118
Qy 119 GIASMP-----PLTAVAPV--PWG-----SIPVGMSSPLVSSVPPAVPLANGCAP 164
Db 120 ---SMPNLISHQPLPVPVPIATPLSSATSGTSLPPLMMPAPLVPSVSTSLP---NGTAS 173

```

```

QY 165 VIOPLAFAPATWPKSSSES--RSGPSQNTKLQKQS--PDVASAP----- 211
DB 174 LIQPL--SIYSSSTLPHASSYSLMGFG--CASIQKASLIDLSSSTSTSTASLSCN 229
QY 212 ----AAEMAVPOSSRLKYROLFNSHDKTMSGLTGPQARTIIMOGSLPQALASIMWS 266
DB 230 SPKGTSEMAVPOPSRLKYQKFNLSLDKMGSGYLSGFGAANALQGNLSQTLATITWLA 289
QY 267 DIQDGLTLEEFFLMLHLDVAMSGOPLPVLPPVIPPSPFRVSGSGMSYSSSSVD 326
DB 290 DIGDGLKKEEFFLMLHLDVAMSGOPLPVLPPVIPPSPFRVSGSGMSYSSSSVD 341
QY 327 QRLPEESSDEDOPEKKLPVTFEDKKRENFERGVSLEKRRQALLQORKEQERLAQE 386
DB 342 GTPSYQKTO--EERPOKKLPVTFEDKKRENFERGVSLEKRRQALLQORKEQERLAQE 400
QY 387 RAEQERKERROEOEAKROULEKOLEKOREROREEERKEITEREAKRLEEROROL 446
DB 401 KEEMERKORERLOEOMWKOLELEKRELEROREEERKEITEREAKRLEEROROL 460
QY 447 EEMERNROELINORNKQOEGTVLKKARKTLEFELEALNDKKQLDGRICRLATOR 506
DB 461 EEMERLRQELLSQKTRQEDIVLSSRKSLHLELAENVKHOQISGRIDVQIRKQTK 520
QY 507 QEISTNKSEBELRIATLHLOOLOSQOMLGRILPEKQILSDQKOVQONSILHRSLLT 566
DB 521 TELEVLDKOCDLEIMEIKOLEKEKONKLIYLVPEKQILNERIKNMQNSNTPDSGISL 580
QY 567 LKRALKEKELAROOERLEVERETSKLOEIDVFNNOLKEIREHSKQOLKORSLA 626
DB 581 LHKSSSEKEELCORLKEQDLAKETASKLSEMDSPNQKLEIRESNTOQLLEQHLKI 640
QY 627 ARLKOEKERSLE--LEKOK---EDAQRVOERDKOMLEHVQOEOPRRPKRHEEDRLK 682
DB 641 KRDKLKEIERLEJOIKKLEDEARKAKOKENLWRESIRKEE-----EKORLOE 694
QY 683 EDSVRKKEAEERAKPEMODOSRLFHHQEPKATATOAPSTTEKGLTISAGESVKVVY 742
DB 695 EKSDKTOEER--KAERKQSE-----TASA-----LVN 721
QY 743 YRALPEESRSHDEITTOQDGIWVWDESQTEGPMLGELKGTGMFPANVAKIPENEY 802
DB 722 YRALPYEARNHDMSSSGDIIQVDEKTVGEPGMLVSGFGKMGPCNYVEKVLSE-- 780
QY 803 PTPAKPVTLTSAAPKALARETAPLPVTSSEPTPNMADPSSSTWBSSEKERTDN 862
DB 781 -----KALSPKKALLPPTVLSATSTST--SOPPAVTDYHNV--SFSMLTYNT-- 824
QY 863 WDTWAAQPSLTVPSAGOLRORSATTPATATGSSSPVLAGGEKEVEGLQAOALYPMRAKD 922
DB 825 --TW-----QKSAFTRIVSPG--SVSPHGGQAVENLKAQALCSMTAKKE 867
QY 923 NMLNFKSDVITVLEODMMWFGVQOKGMPKSYKYLISG--PYRKS--TSIDTGTES 979
DB 868 NMLNFKSHDVTIVLEOQENMWFGEVHGGRGMPKSYKYLIPGNEVORGEDEALVAAVTKK 927
QY 980 PASLKRVAAPAKAIP--GEFFIMATYEESEOGDLTFQOQDVIWVTKXGDMWTGVGD 1038
DB 928 PTS---TAYPYTATYAPVGEYIALYISYSVEPBDLFTTEGEBILVYQKGEWMTSGICE 964
QY 1039 KSGVFPNSNYRLKDSGSGTAGTKSLGKKPEIAQVIASVATGPEQLTAPQOLITRK 1098
DB 985 RTGIFPSNYVAPKQOENGNASKSGANSKREIAQVITSAVAGSTEQSLAPQOLITLK 1044
QY 1099 KNPBGWMEGELQARQKQIQGMFPANVYKLLSPGSKITPPELPTKAVOAVOQVGMVD 1158
DB 1045 KNTSGMWQEGELQARQKQIQGMFPANVYKLLSPGSKITPPELPTKAVOAVOQVGMVD 1097
QY 1159 YTAQNDDELAFSGQIIVLNKEDPDMWKGEVSGQVGLFSPSNYKLLTTDMPSQ 1213
DB 1098 YMANNEDELINSGQIIVLNKEDPDMWKGEVSGQVGLFSPSNYKLLTTDMPSQ 1152

```

```

RESULT 14
AA71160
ID AA71160 standard; Protein; 1683 AA.
AC AA71160;
DT 08-SEP-2000 (first entry)
DE
XX Rat phosphodiesterase interacting protein, M14.
XX
XX Rat; phosphodiesterase interacting protein; M14; PDE; CAMP-PDE;
XX cyclic adenosine monophosphate phosphodiesterase; asthmal;
XX antiinflammatory; antipruritic; dermatological; antibacterial; shock;
XX analgesic; immunosuppressive; antitumor; vasodilator; anticholinergic;
XX antidiabetic; urticaria; antiallergic; antiatherosclerotic; diagnosis;
XX antineoplastic; treatment; inflammatory disease; psoriasis; arthritis;
XX atherosclerosis; Crohn's disease; cystic fibrosis; chronic bronchitis;
XX eosinophilic granuloma; proliferative skin disease; ulcerative colitis;
XX reperfusion injury; atopic dermatitis; diabetes insipidus;
XX conjunctivitis; adult respiratory distress syndrome; allergic rhinitis;
XX arterial restenosis; ankylosing spondylitis; transplant rejection;
XX graft versus host disease.
XX
XX Rattus sp.
XX
XX WO200027861-A1.
XX
XX 18-MAY-2000.
XX
XX 12-NOV-1999; 99MO-US6860.
XX
XX 12-NOV-1998; 98US-010825.
XX
XX (STRD ) UNIV LELAND STANFORD JUNIOR.
XX
XX Conti M, Pahlke G;
XX
XX WPI; 2000-376479/32.
XX
XX Polynucleotide encoding a phosphodiesterase (PDE) interacting
XX PT polypeptide, useful for diagnosis and treatment of asthma, cystic
XX PT fibrosis, Crohn's disease, and rheumatoid arthritis -
XX
XX PS Disclosure; Fig 6; 77pp; English.
XX
XX The present sequence is a phosphodiesterase (PDE) interacting protein,
XX M14 from rat. The protein modulates the functions and properties of PDEs,
XX specifically CAMP-PDEs, and also targets them to specific subcellular
XX CC compartments. The present sequence
XX CC associated with PDE activity. The diseases include asthma, cystic
XX CC fibrosis, inflammatory airway disease, chronic bronchitis, eosinophilic
XX CC granuloma, psoriasis, proliferative skin diseases, endotoxin shock,
XX CC septic shock, ulcerative colitis, Crohn's disease, reperfusion injury,
XX CC inflammatory arthritis, atopic dermatitis, urticaria, adult respiratory
XX CC distress syndrome, diabetes insipidus, allergic rhinitis, allergic
XX CC conjunctivitis, vernal conjunctivitis, arterial restenosis,
XX CC atherosclerosis, inflammatory diseases associated with irritation and
XX CC pain, rheumatoid arthritis, ankylosing spondylitis, transplant
XX CC rejection and graft versus host disease, disease conditions associated
XX CC with hypersecretion of gastric acid, and disease conditions in which
XX CC cytokines are mediators.
XX
XX Sequence 1683 AA;
XX
XX Query Match 48.6%; Score 3046; DB 21; Length 1683;
XX Best Local Similarity 50.8%; Pred. No. 9,4e-187;
XX Matches 652; Conservative 187; Mismatches 267; Indels 178; Gaps 33;
QY 1 MAQFPTFGSGLDVVAITVEERAKHDOFLSLKPIAGFITGDOARNFPGGLPQVLAQ 60
DB 2 MAQFPTANMGGNMMAITSEERTKHDKQFDNLKPSGSGYITGDOARTFLOGLPAPVLA 61

```

```

QY 61 IVALADNMNDGRMDQVEFSIAMKLILKLOGYOLPSTLPPWMOQV--AISSAPFGIG 118
DB 62 IVALADNMNDGRMDQVEFSIAMKLILKLOGYOLPSTLPPWMOQV--AISSAPFGIG 120
QY 119 GIASMP-----PLTAAVAV--PMG-----SIPVGMSPPLVSSPPAAVPLANGAP 164
DB 121 ---SMRPLSHOQLPPVAPITAPLSATSTGTSIPPLMAPPLVPSVSTSSLP---NGTAS 174
QY 165 VIQPLPAFAHAPATWPKSSFS--RSGPGSOLNTKLOKAGS--FDVASAPP----- 211
DB 175 LIQPL-SIPYSSSTLPHASSYSIMMGFG--GASIQKAGSLIDLSSSTSTSTASLJSGN 230
QY 212 -----AAEMAVPOSSRLKRYQLFNSHDKTMSGHLTGPQARTIMOSLSLOAQASLWINS 266
DB 231 SPRTGSEMAVVPQPSRLKRYQKFNSLDKSMSGYLSGFOANALLOSLOTQATITWTA 290
QY 267 DIQDQKLTAEETILAMHLIDVAMSGOPLPPVLPPEYIPSPFRVSGSGMSVSSSSVD 326
DB 291 DIQDQKLTAEETILAMHLIDVAMSGOPLPPVLPPEYIPSPFRVSGSGMSVSSSSVD 342
QY 327 QRLPEEPPSEDEOQPEKKLVTFEDKKRENFERGSAVELEKRRQALLQOQKEQERLAQLE 386
DB 343 GTLPYSYQKTO-ESEPQKLPVTFEDKKRKNYERGNMELERQVLMEQOQREAKQAQKE 401
QY 387 RAQOEKREKROQEAQKOLELEKOLEKOREKOREBERKKEIFEREAQKREKQOL 446
DB 402 KEEMERKORELOEQEMKOLELEKOLEKOREKOREBERKKEIFEREAQKREKQOL 461
QY 447 EMERNRROELNDRNKEOEGTVVLKARKTLEFELALNDKHKQLEGLKODIRCLATQR 506
DB 462 EMERIRROELNDRNKEOEGTVVLKARKTLEFELALNDKHKQLEGLKODIRCLATQR 521
QY 507 QEISTNKSRELIAETILHLOQLOESQOMLGRLEPEKOLISDOLKOVONSLHRDLSLT 566
DB 522 TELEVDKQCDLIMEIKOQLOQLOEVONKLIYLVPEKOLNRIKQMLNSTPDSGJL 581
QY 567 LKALAEKELARQOLREOLDEVERETRSKLOETIDVFNNO----- 605
DB 582 LHKSSSEKELCQRLKEQDLEKELTASKEEMDSFNQOLKCGNMDSVLOCLLSLJLCL 641
QY 606 -----LKELEHSHKQOOLOKORSLAEARKLOKEOERKSLF-LEKOK--EDAQRVQER 655
DB 642 NMLFLLEKELREBYNTQOLAEQHLKIKDKLELEKREKLEKQKLEDEAKRKAQKQK 701
QY 656 DKQWLEHVOEEOEPKRPKHEEDRLKREDSVRKKEAEERAKPEMOKOSRLPHQOEPK 715
DB 702 ENLMKRSIRKEE-----EKQKRLQEKSOQDRTQEBE----- 734
QY 716 LATQAFMSTTEKPLTISAQESVAVVYRALYPPESRSHDEITQPGDIVMDESQTEGP 775
DB 735 -KTEAKQSESTARA-----LVNYRALYFEARNDHEMSFNGDIIQVDEKTVGEP 782
QY 776 GWLGGELKGTGWFAPNVAEK-IPENEVTPRAK---PVTDLISAPKALARETAPLP 830
DB 783 GWLYGSPQGFQFPCNVEKMLSSKDTBPKALLPRAVLSATSA-----APQ 833
QY 831 VTSSEBPTTPNNMADFSSTWSSSNEKPEPTMDTAAQPSLTVPSAGOLRQSAFTPAT 890
DB 834 LGSNGAPAV---TDYQNV--SFSNLAVNT---TW-----QOKSATRTV 869
QY 891 ATGSSSPSVLQGEKEVGLQAOALYPMRAKKDNLNFKSDVITVLEQODMMWFGEVQO 950
DB 870 SPG-SVSPHIGQOAVENLKAQALCSWTAKENHLNFSKHDVITVLEQOENWFGVHG 928
QY 951 KGFPPSYVYLISGPAVKSTISDTGPTES--PASLKRVASPAAPALPGEFTAMTYEES 1009
DB 929 KGFPPSYVYLISGPAVKSTISDTGPTES--PASLKRVASPAAPALPGEFTAMTYEES 981
QY 1010 EOODLTFQOQDVLIVTKDDMMTGTVDGKSGVFPNNYVRLKDSGSGTAGTSGIGKPP 1069
DB 982 EPDPLTFTEBEELLVYQKDSGEMWTSGIGERTGTFPSNYVRPKQDEVANGASGSAKNKP 1041
QY 1070 EIAQVIASVATGPBQULTAPGQLILIRKKNGPGWEGELQAGKKRQIGWFPANYVLL 1129

```

```

DB 1042 EIAQVTSVYVAAAGAEQSLASGOLILILKXSSSGMCGEIQARERKQKGFPPASHVLL 1101
QY 1130 SPGTSKIPTELPATAVQAVQVIGWYDYTAQNDDELATFSKQIINVLNKEDPDMMKGE 1189
DB 1102 GPSARTTP-----AFHACVQIYAMTDYIANNDELNFSKGLINVMKDDPDMMQGE 1154
QY 1190 VSGQVGLFPPSNVYKLTDTMDPSQ 1213
DB 1155 INGVGLFPPSNVYKMTTDSQ 1178

RESULT 15
AAM93229
ID AAM93229 standard; Protein; 642 AA.
AC AAM93229;
XX
XX
DT 06-NOV-2001 (first entry)
XX
XX
DE Human polypeptide, SEQ ID NO: 2647.
KW Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
XX
OS Homo sapiens.
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
XX
PA (HELI-) HELIX RES INST.
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y,
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H,
DR WPI: 2001-524255/58.
DR N-PSDB; AAK94139.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Claim 8; SEQ ID NO 2647; 1380bp + sequence listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesized by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a polypeptide
CC encoded by a full length human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 642 AA;

Query Match 48.4%; Score 3034.5; DB 22; Length 642;
Best Local Similarity 93.9%; Pred. No. 1,3e-186;
Matches 603; Conservative 18; Mismatches 20; Indels 1; Gaps 1;

QY 1 MAOFPPTGGSLDWATLVEERAKGDOQFLSLKPIAGFITGDOARNFPOSGLPQVLAQ 60
DB 1 MAOFPPTGGSLDWATLVEERAKGDOQFLSLKPIAGFITGDOARNFPOSGLPQVLAQ 60
QY 61 IVALADNMNDGRMDQVEFSIAMKLILKLOGYOLPSTLPPWMOQVPAISSAPFGIGI 120

```



```

Db 61 IWLADANNNOQVDFESIAMKLIKLOQYOLPSALPVMKQOPPAISSAPAFGWI 120
QY 121 ASMPPLTAVAPVPMGSI:PVVGMSPLVSVPPAVPPLANGAPPVIOPLPAFAHPATWP 180
Db 121 ASMPPLTAVAPVPMGSI:PVVGMSPLVSVPPAVPPLANGAPPVIOPLPAFAHPATWP 180
QY 181 KSSSFSRSGSQSUNTLOKAQSFVASAPPAEMAVPOSSRLKYROLFNSHDKTMSGHL 240
Db 181 KSSSFSRSGSQSUNTLOKAQSFVASAPPAEMAVPOSSRLKYROLFNSHDKTMSGHL 240
QY 241 TGPOARTILMOSSL:POAQLASIMWLSDIDODGKLTAEFFILAMHLIDVAMSGQPLPVLP 300
Db 241 TGPOARTILMOSSL:POAQLASIMWLSDIDODGKLTAEFFILAMHLIDVAMSGQPLPVLP 300
QY 301 PEYIPPSFRVRSGSGSVISSSVSDORLPPEPSSDEQOP-EKKLPVTFEDKKRENER 359
Db 301 PEYIPPSFRVRSGSGSVISSSVSDORLPPEPSSDEQOP-EKKLPVTFEDKKRENER 359
QY 360 GSVELKKRQALBEOQRKEQERLAQLERABOERKEREQOEAQRQOLEKOLEKORELE 419
Db 361 GNLLEKKRQALBEOQRKEQERLAQLERABOERKEREQOEAQRQOLEKOLEKORELE 420
QY 420 ROEEERKEIEREAKRELERQOLEWERNROELLNOANKOEGTVLAKARKTLEF 479
Db 421 ROEEERKEIEREAKRELERQOLEWERNROELLNOANKOEGTVLAKARKTLEF 480
QY 480 ELBALNDKHOLEGKLODIRCRLATOROEIESTNKSRELRIAEITHLOOQLOESQOMLGR 539
Db 481 ELBALNDKHOLEGKLODIRCRLATOROEIESTNKSRELRIAEITHLOOQLOESQOMLGR 540
QY 540 LIPEKOIISDOLKOYQONS:LRDSLLTKRALAKELARQOLREQLDVEVERETRSKLOEI 599
Db 541 LIPEKOIISDOLKOYQONS:LRDSLLTKRALAKELARQOLREQLDVEVERETRSKLOEI 600
QY 600 DVFNNOLEKEIRISKOOLOKORSLEAARLKOEQERKSLEL 641
Db 601 DVFNNOLEKEIRISKOOLOKORSLEAARLKOEQERKSLEL 642

```

Search completed: March 14, 2003, 11:37:45  
 Job time : 79 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2003, 11:37:26 ; Search time 476 Seconds  
(without alignments)  
117.457 Million cell updates/sec

Title: US-09-674-237A-3

Perfect score: 6263  
Sequence: 1 MAQFPTPGSGSDVMAITVE.....VGLFSPSNVYKLTMDPSQO 1213

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 199416 seqs, 46092074 residues

Total number of hits satisfying chosen parameters: 199416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2058.5	32.8	462	10 US-09-879-957-38	Sequence 38, Appl
2	1388.5	22.1	509	10 US-09-879-957-194	Sequence 194, App
3	1330.5	21.2	530	9 US-09-764-868-738	Sequence 738, App
4	816.5	13.0	248	10 US-09-879-957-40	Sequence 40, Appl
5	527	8.4	813	9 US-09-864-899-25	Sequence 25, Appl
6	429	6.8	90	10 US-09-864-761-33620	Sequence 33620, A
7	429	6.8	90	10 US-09-864-761-43004	Sequence 43004, A
8	426	6.7	99	10 US-09-864-761-31635	Sequence 31635, A
9	422	6.7	82	10 US-09-864-761-46565	Sequence 46565, A
10	413	6.6	87	10 US-09-864-761-34169	Sequence 34169, A
11	325	5.2	659	9 US-10-144-621-2	Sequence 2, Appl
12	322	5.1	59	10 US-09-879-957-133	Sequence 133, App
13	316.5	5.0	287	9 US-09-764-868-744	Sequence 744, App
14	314	5.0	62	10 US-09-879-957-135	Sequence 135, App
15	309	4.9	56	10 US-09-879-957-114	Sequence 114, App
16	303.5	4.8	611	10 US-09-216-393-81	Sequence 81, Appl
17	300.5	4.8	1175	10 US-09-771-161A-224	Sequence 224, App
18	300.5	4.8	1175	10 US-09-771-161A-225	Sequence 225, App
19	300.5	4.8	1175	10 US-09-771-161A-226	Sequence 226, App

20	296	4.7	57	10 US-09-879-957-136	Sequence 136, App
21	293	4.7	88	9 US-09-893-519A-73	Sequence 73, Appl
22	288	4.6	59	10 US-09-864-761-33618	Sequence 33618, A
23	287.5	4.6	76	10 US-09-864-761-34168	Sequence 34168, A
24	287.5	4.6	82	10 US-09-864-761-46614	Sequence 46614, A
25	287.5	4.6	150	9 US-10-101-487-58	Sequence 58, Appl
26	285.5	4.6	830	9 US-10-033-245-7	Sequence 7, Appl
27	285.5	4.6	830	9 US-10-033-223-7	Sequence 7, Appl
28	285.5	4.6	830	9 US-10-033-167-7	Sequence 7, Appl
29	285.5	4.6	830	9 US-10-033-244-7	Sequence 7, Appl
30	285.5	4.6	830	9 US-10-033-435-7	Sequence 7, Appl
31	285.5	4.6	830	9 US-10-033-990-7	Sequence 7, Appl
32	285.5	4.6	830	12 US-10-033-246-7	Sequence 7, Appl
33	285.5	4.6	830	12 US-10-033-301-7	Sequence 7, Appl
34	285.5	4.6	830	12 US-10-033-326-7	Sequence 7, Appl
35	281	4.5	788	10 US-09-879-957-30	Sequence 30, Appl
36	277	4.4	97	10 US-09-864-761-36740	Sequence 36740, A
37	277	4.4	101	10 US-09-864-761-37838	Sequence 37838, A
38	277	4.4	103	10 US-09-864-761-36224	Sequence 36224, A
39	275	4.4	1360	10 US-09-871-916-2	Sequence 2, Appl
40	273.5	4.4	1005	10 US-09-925-301-1335	Sequence 1335, Ap
41	271	4.3	88	10 US-09-864-761-36232	Sequence 36232, A
42	270.5	4.3	2568	10 US-09-866-108-3	Sequence 3, Appl
43	264	4.2	1805	9 US-09-820-843A-73	Sequence 73, Appl
44	264	4.2	1945	9 US-09-927-597-2	Sequence 2, Appl
45	264	4.2	1979	9 US-09-927-597-4	Sequence 4, Appl

#### ALIGNMENTS

RESULT 1  
US-09-879-957-38  
Sequence 38, Application US/09879957

Patent No. US20020034755A1

GENERAL INFORMATION:

APPLICANT: SPARKS, Andrew B.

HOFFMAN, No. US20020034755A1h

KAY, Brian K.

FOWLER, Dana M.

MCCONNELL, Stephen J.

TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL

DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND

USING SAME

NUMBER OF SEQUENCES: 227

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/879, 957

FILING DATE: 13-Jun-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/630, 915

FILING DATE: 03-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: Mistic, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 1101-174

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864/9741

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 38:



RESULT 3  
 US-09-764-868-738  
 ; Sequence 738, Application US/09764868  
 ; Patent No. US2002016871A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: PT232  
 ; CURRENT APPLICATION NUMBER: US/09/764,868  
 ; CURRENT FILING DATE: 2001-01-17  
 ; Prior application data removed - refer to PALM or file wrapper  
 ; NUMBER OF SEQ ID NOS: 1510  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 738  
 ; LENGTH: 530  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-09-764-868-738

Query Match 21.2%; Score 1330.5; DB 9; Length 530;  
 Best Local Similarity 49.9%; Pred. No. 4.1e-57;  
 Matches 269; Conservative 80; Mismatches 115; Indels 75; Gaps 12;

QY 679 RLKREDSVRKKEAEERAKPEMOKOSRLFHHPQEPKALATQAPWSTTEKGPLTISQESV 738  
 DB 4 RRTKATPGKKTKQEK-----IOEERKAEKQRETLASV----- 37  
 QY 739 KVVYRYALYPEFSRSHDEITIQGDIYVWDESOTGEFGWLGELKGTGFPANYAEKIP 798  
 DB 38 -LVNYRYALYPEFARNHDEMSFNSGDIIOVEKTVGEGWLYGSGFQNGFPCNYEKRP 96  
 QY 799 -EENEVTPAKPVDTLTSAPAPKALRETPAPLPVTSSEPTTPNNMADFSSSTPSSE 856  
 DB 97 SSENK-----KAVSPKKAL-----LPPTVLSATS-----TSSEPLSSNQ 131  
 QY 857 KPETDMDWTMAAGPSLTVPSAGQLRQSAFTPATGSSSPVLGGEKYEGLQALALP 916  
 DB 132 PASVTYDQN-VSFSNLTIVNSWQ--KKSATRTVSPG-SVSPTHGGQGVVENKKAQALCS 187  
 QY 917 WRAKKHNLHFNKSDVITVLEQDDMMWFGVQOKGWFPSYVKLISGPRKSTSIDTGP 976  
 DB 188 WTKAKDNHLNFSKHDITITVEQDENMWFGEVHGRCWFPSYVKIIPGSEVK----- 239  
 QY 977 TESPASIKRVAS--PAKPAIPGEETAMTYESSGQDLTFQOQGVIVVTKKGDPMWTG 1034  
 DB 240 REEPALYAIVNKKPTSAAYSGEYIALPYSGVEPGLTFTEGELLVITQDGEWMTG 299  
 QY 1035 TVGDKSGVFNPNYRLKDSGSGTAGTSLGKKPELAQVIASATGPEQLTLAAGQLI 1094  
 DB 300 SIDRSGIIFPSNVKPKDQESFGSASKSGSNKKPELAQVTSAYVASGSQSLAPGOLI 359  
 QY 1095 LIRKKNPGWMEGELQARGKKROIIGFPPANYVKLISFGTSKITPELPTAVCPAVCOVI 1154  
 DB 360 LILKKTSGWMEGELQARGKKROIIGFPPASHVKLLGPSSSRATPAFHP-----VCQVI 412  
 QY 1155 GWYDYTAQNDELAFSGKQIINVLNKEDPDWKGEGVGQVGLFPPSNYVKLITDMDPSQ 1213  
 DB 413 AMYDYANNDELFSKQILINWKNKDDPDWMEGELNGVGLFPPSNYVKLITDSDPSQ 471

RESULT 4  
 US-09-879-957-40  
 ; Sequence 40, Application US/09879957  
 ; Patent No. US20020034755A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: SPARKS, Andrew B.  
 ; HOFFMAN, No. US20020034755A1  
 ; KAY, Brian K.  
 ; FOMLICKS, Dana M.  
 ; MCCONNELL, Stephen J.  
 ; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL  
 ; DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
 ; USING SAME

NUMBER OF SEQUENCES: 227  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Pennie & Edmonds LLP  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10036-2711  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/879,957  
 FILING DATE: 13-Jun-2001  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/630,915  
 FILING DATE: 03-APR-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Mistrock, S. Leslie  
 REGISTRATION NUMBER: 18,872  
 REFERENCE/DOCKET NUMBER: 1101-174  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 790-9090  
 TELEFAX: (212) 869-8864/9741  
 TELEX: 66141 PENNIE  
 INFORMATION FOR SEQ ID NO: 40:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 248 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: <Unknown>  
 TOPOLOGY: unknown  
 MOLECULE TYPE: peptide  
 SEQUENCE DESCRIPTION: SEQ ID NO: 40:  
 US-09-879-957-40

Query Match 13.0%; Score 816.5; DB 10; Length 248;  
 Best Local Similarity 58.4%; Pred. No. 9.6e-33;  
 Matches 153; Conservative 38; Mismatches 54; Indels 17; Gaps 3;

QY 954 FPKSYVKLISGPRKSTSIDTGPESASLKRVAS--PAKPAIPGEETAMTYESSSEQ 1011  
 DB 2 FFAASTKILIPSEVK-----REEPALYAIVNKKPTSAAYSGEYIALPYSSVEP 53  
 QY 1012 GDLTFOQGVIVVTKKGDPMWTGTVGDKSGVFNPNYRLKDSGSGTAGTSLGKKPEI 1071  
 DB 54 GDLTTEBEELITVQKGEWMTGSIGRSGIFPSNVKPKDQESFGSASKSGASNKKPEI 113  
 QY 1072 AQVLSAYATGPEQLTAPGQLILIRKKNPGWMEGELQARGKKROIIGFPPANYVKLLSP 1131  
 DB 114 AQVLSAYASGSEQLAPGQLILILKKNPSGWMEGELQARGKKROIIGFPPASHVKLLGP 1173  
 QY 1132 GTSKITPELPTAVCPAVCOVIGYDYTAQNDELAFSGKQIINVLNKEDPDWKGEGVS 1191  
 DB 174 SSERATPAFHP-----VCQVIAMYDYANNDELFSKQILINWKNKDDPDWMEGELIN 226  
 QY 1192 GQVGLFPPSNYVKLITDMDPSQ 1213  
 DB 227 GVTGLFPPSNYVKMTTSDPSQ 248

RESULT 5  
 US-09-964-899-25  
 ; Sequence 25, Application US/09964899  
 ; Patent No. US2002017446A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cohen, Dalja et al.  
 ; TITLE OF INVENTION: Identification of Genes Involved in  
 ; TITLE OF INVENTION: Alzheimer's Disease Using Drosophila Melanogaster  
 ; FILE REFERENCE: 4-31612 A  
 ; CURRENT APPLICATION NUMBER: US/09/964,899



OTHER INFORMATION: EST\_HUMAN HIT: AA773823.1, EVALU2.00e-37  
 OTHER INFORMATION: SWISSPROT HIT: Q15811, EVALU2.00e-45  
 US-09-864-761-33620

Query Match 6.8% Score 429; DB 10; Length 90;  
 Best Local Similarity 94.0%; Pred. No. 1.4e-14;  
 Matches 79; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 767 VDESQTGEPCWLGELKGTGWPFPANYAEKIPENEVPTPAKPYTDLTSAPAPKALRETP 826

DB 1 VDESQTGEPCWLGELKGTGWPFPANYAEKIPENEVPTPAKPYTDLTSAPAPKALRETP 60

QY 827 APLPVTSSPSTTPNNWADFSSTW 850

DB 61 APLAVTSSPSTTPNNWADFSSTW 84

RESULT 7

US-09-864-761-43004  
 Sequence 43004, Application US/09864761  
 Patent No. US20020048763A1  
 GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 FILE REFERENCE: Aeonica-X-1

CURRENT FILING DATE: 2001-05-23

PRIOR FILING DATE: 2001-05-23

PRIOR FILING DATE: 2000-02-04

PRIOR FILING DATE: 2000-02-04

PRIOR FILING DATE: 2000-05-26

PRIOR FILING DATE: 2000-08-03

PRIOR FILING DATE: 2000-10-04

PRIOR FILING DATE: 2000-09-27

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

FEATURE:

OTHER INFORMATION: MAP TO AP000311.1

OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.87

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1

OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3

OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 32

OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.66

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.69

OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1

OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.92

OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.98

OTHER INFORMATION: EST\_HUMAN HIT: AA773823.1, EVALU2.00e-37

OTHER INFORMATION: SWISSPROT HIT: Q15811, EVALU2.00e-45

US-09-864-761-43004

Query Match 6.8% Score 429; DB 10; Length 90;  
 Best Local Similarity 94.0%; Pred. No. 1.4e-14;  
 Matches 79; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 767 VDESQTGEPCWLGELKGTGWPFPANYAEKIPENEVPTPAKPYTDLTSAPAPKALRETP 826

DB 1 VDESQTGEPCWLGELKGTGWPFPANYAEKIPENEVPTPAKPYTDLTSAPAPKALRETP 60

QY 827 APLPVTSSPSTTPNNWADFSSTW 850

DB 61 APLAVTSSPSTTPNNWADFSSTW 84

RESULT 8

US-09-864-761-33635  
 Sequence 33635, Application US/09864761  
 Patent No. US20020048763A1  
 GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
 FILE REFERENCE: Aeonica-X-1

CURRENT FILING DATE: 2001-05-23

PRIOR FILING DATE: 2001-05-23

PRIOR FILING DATE: 2000-02-04

PRIOR FILING DATE: 2000-02-04

PRIOR FILING DATE: 2000-05-26

PRIOR FILING DATE: 2000-08-03

PRIOR FILING DATE: 2000-10-04

PRIOR FILING DATE: 2000-09-27

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

PRIOR FILING DATE: 2001-01-30

```

; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33635
; LENGTH: 99
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000311.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2
; OTHER INFORMATION: SWISSPROT HIT: 035601, EVALUATE 2.80e-01
; OTHER INFORMATION: EST_HUMAN HIT: AA815076.1, EVALUATE 1.00e-08
; US-09-864-761-33635
```

```

Query Match          6.8%; Score 426; DB 10; Length 99;
Best Local Similarity 93.5%; Pred. No. 2.1e-14;
Matches 86; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
```

```

QY 348 TFEKKRENERGVLEKRRQALLEGORKEQERLQLEBAERKREKREOEAKROLE 407
      |||||.....|.....|.....|.....|.....|.....|.....|
DB 1 TFEKKRENERGVLEKRRQALLEGORKEQERLQLEBAERKREKREOEAKROLE 60
```

```

QY 408 LEKQLEKQELERQREERREKRETERREAKRE 439
      |||||.....|.....|.....|.....|.....|.....|.....|
DB 61 LEKQLEKQELERQREERREKRETERREAKRE 92
```

```

RESULT 9
US-09-864-761-46565
; Sequence 46565, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
```

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 46565
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000117.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.76
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95
; OTHER INFORMATION: SWISSPROT HIT: 015811, EVALUATE 9.00e-45
; OTHER INFORMATION: EST_HUMAN HIT: AA73823.1, EVALUATE 2.00e-37
; US-09-864-761-46565
```

```

Query Match          6.7%; Score 422; DB 10; Length 82;
Best Local Similarity 95.1%; Pred. No. 2.7e-14;
Matches 78; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```

QY 767 VDESOTGEPGWLGGELKKTGTGFNPNYAETKIPENVEPPPAKPVDTLSAPAPKTLARETP 826
      |||||.....|.....|.....|.....|.....|.....|.....|
DB 1 VDESOTGEPGWLGGELKKTGTGFNPNYAETKIPENVEPPPAKPVDTLSAPAPKTLARETP 60
```

```

QY 827 APLVTSSEPTTPNNMADFS 848
      |||||.....|.....|.....|.....|.....|.....|.....|
DB 61 APLVTSSEPTTPNNMADFS 82
```

```

RESULT 10
US-09-864-761-34169
; Sequence 34169, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/00665
```



PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 34169  
LENGTH: 87  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AP000049.1  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.3  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.3  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3  
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.9  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.2  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.5  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.3  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9  
OTHER INFORMATION: EST HUMAN HIT: AA815076.1, EVALUATE 4.00e-09  
OTHER INFORMATION: SWISSPROT HIT: Q35601, EVALUATE 1.80e-01  
US-09-864-761-34169  
Query Match  
Best Local Similarity 96.6%; Score 413; DB 10; Length 87;  
Pred. No. 7.9e-14;  
Matches 84; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 348 TFFDDKRENERGSEVLEKRRQALLLEQKREQLAQLERAEERKEREQDEAKRQLE 407  
DB 1 TFFDDKRENERGSEVLEKRRQALLLEQKREQLAQLERAEERKEREQDEAKRQLE 60  
QY 408 LEKQLEKQRELEKQREERKEREIERRE 434  
DB 61 LEKQLEKQRELEKQREERKEREIERRE 87  
RESULT 11  
US-10-144-621-2  
Sequence 2, Application US/10144621  
Patent No. US20020172671A1  
GENERAL INFORMATION:  
APPLICANT: Hsu, Yen-Ming  
TITLE OF INVENTION: THE CAIP-LIKE GENE FAMILY  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street

CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows 95  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/144,621  
FILING DATE: 13-May-2002  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/671,354  
FILING DATE: 27-JUN-1996  
APPLICATION NUMBER: US 08/486,344  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/484,709  
FILING DATE: 07-JUN-1995  
APPLICATION NUMBER: US 08/475,894  
FILING DATE: 07-JAN-1995  
APPLICATION NUMBER: US 08/475,710  
FILING DATE: 07-JAN-1995  
APPLICATION NUMBER: US 08/474,697  
FILING DATE: 07-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Myers, Louis  
REGISTRATION NUMBER: 35,965  
REFERENCE/DOCKET NUMBER: 10274/009005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 659 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
FEATURE:  
LOCATION: 29...29  
OTHER INFORMATION: where Xaa at position 29 is Arg, Cys, Gly, or Ser  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-10-144-621-2  
Query Match  
Best Local Similarity 5.2%; Score 325; DB 9; Length 659;  
Pred. No. 1.2e-08;  
Matches 91; Conservative 67; Mismatches 115; Indels 64; Gaps 15;  
QY 914 LYPMAKKDNLHFNPKSVITVLEQD-MMMFGEVQQKGFPSYVKLISGPRKSTSI 972  
DB 1 IYVYLRHDXDEXTISVGIITNIREKDGXWEGQINGRGLFPNFAREIKKEXKXDPDT 60  
QY 973 DTGPTSEASIKRYVSPAKPAIPGEFEI-----AMTYESESQGLDFQO 1018  
DB 61 NKAP-EKP--LHEV--PGNSLLSSETLRTNKKGERRRRCQVAFSYLPGNDELKLV 115  
QY 1019 GDVI-VVTKKDGDMWTGTVDGKSGVFPNSNYR-----LKDSGS 1056  
DB 116 GDILEVGEVEEGWEGVLTGKTCFSPNFIKELSGSDDELGISQDBQKSKSLRETTGS 175  
QY 1057 GTAGKTSGLKKPELQVIAVYATGPOL-----TLAPQQLLIRKKPNKGWGELEQ 1110  
DB 176 ESDGSDS-STKSGANGTAVTAIIPKRVGVGFQDIPDKPIKLPRRS---IEVEND 230  
QY 1111 ARGKRGIG-WFPANVYVLSPGTSKITPTLPTAVQPAVCQYIGWYDYTAQNDDELAF 1169  
DB 231 FLVPEKTKGKLPAF---TATPDSK---TEMDSRTSKXQYCKYI--FPYEAQNDDELTI 282  
QY 1170 SKGQILNVLNKE--DPDWMKGEVSGQVGLFPSNYKL 1204

Db 283 KEGDIVTLINKCIDVGMWEGELNGRGVFPDNFVKL 319

## RESULT 12

US-09-879-957-133

Sequence 133, Application US/09879957

Patent No. US20020034755A1

GENERAL INFORMATION:

APPLICANT: SPARKS, Andrew B.

HOFFMAN, No. US20020034755A1h

KAY, Brian K.

FOWLES, Dana M.

MCCONNELL, Stephen J.

TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL

DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND

USING SAME

NUMBER OF SEQUENCES: 227

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/879,957

FILING DATE: 13-Jun-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/630,915

FILING DATE: 03-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: Mistrock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 1101-174

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864/9741

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 133:

SEQUENCE CHARACTERISTICS:

LENGTH: 59 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

MOLECULE TYPE: unknown

TOPOLOGY: unknown

SEQUENCE DESCRIPTION: SEQ ID NO: 133:

US-09-879-957-133

Query Match

Best Local Similarity 5.1%; Score 322; DB 10; Length 59;

Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 KVVVYRALYPESHSHDEITITQPGDIVVWDSQTEGPGWLGELKGTGFPANVYAEKI 59

US-09-879-957-133

Query Match

Best Local Similarity 5.1%; Score 322; DB 10; Length 59;

Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 KVVVYRALYPESHSHDEITITQPGDIVVWDSQTEGPGWLGELKGTGFPANVYAEKI 59

US-09-879-957-133

Query Match

Best Local Similarity 5.1%; Score 322; DB 10; Length 59;

Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 KVVVYRALYPESHSHDEITITQPGDIVVWDSQTEGPGWLGELKGTGFPANVYAEKI 59

US-09-879-957-133

Query Match

Best Local Similarity 5.1%; Score 322; DB 10; Length 59;

Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 KVVVYRALYPESHSHDEITITQPGDIVVWDSQTEGPGWLGELKGTGFPANVYAEKI 59

US-09-879-957-133

Query Match

Best Local Similarity 5.1%; Score 322; DB 10; Length 59;

Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 KVVVYRALYPESHSHDEITITQPGDIVVWDSQTEGPGWLGELKGTGFPANVYAEKI 59

US-09-879-957-133

Query Match

Best Local Similarity 5.1%; Score 322; DB 10; Length 59;

Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 KVVVYRALYPESHSHDEITITQPGDIVVWDSQTEGPGWLGELKGTGFPANVYAEKI 59

US-09-879-957-133

Query Match

Best Local Similarity 5.1%; Score 322; DB 10; Length 59;

Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 KVVVYRALYPESHSHDEITITQPGDIVVWDSQTEGPGWLGELKGTGFPANVYAEKI 59

US-09-879-957-133

Prior application data removed - refer to PALM or file wrapper

NUMBER OF SEQ ID NOS: 1510

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 744

LENGTH: 287

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (146)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (247)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (252)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (252)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (252)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (252)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (252)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (252)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (252)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (252)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (252)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (252)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (252)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (252)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (252)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (252)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (252)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (252)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (252)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (252)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (252)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (252)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (252)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (252)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (252)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (252)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (252)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/879,957  
FILING DATE: 13-Jun-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/630,915  
FILING DATE: 03-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-174  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 135:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 62 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 135:  
US-09-879-957-135

Query Match 5.0%; Score 314; DB 10; Length 62;  
Best Local Similarity 98.3%; Pred. No. 3.2e-09;  
Matches 59; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1070 EIAQVIASVATGPEQLTAPGQILIRKKNPGWMEGELQARGKKRQIGWFPANVYKLL 1129  
Db 3 EIAQVIASVATGPEQLTAPGQILIRKKNPGWMEGELQARGKKRQIGWFPANVYKLL 62

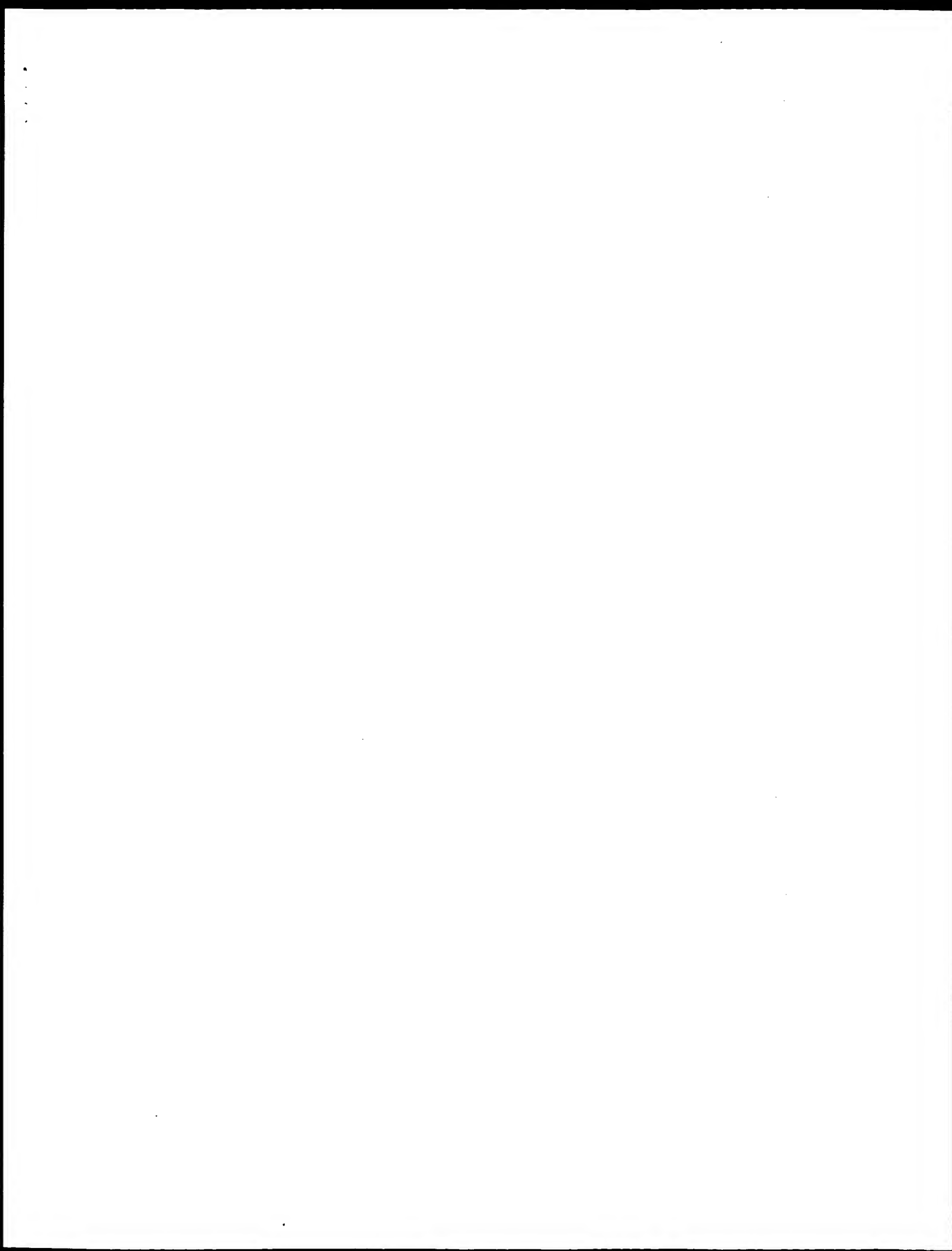
RESULT 15  
US-09-879-957-134  
Sequence 134: Application US/09879957  
Patent No. US20020034755A1  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
HOFFMAN, No. US20020034755A1h  
KAY, Brian K.  
FOWLES, Dana M.  
McCONNELL, Stephen J.  
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL  
DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
USING SAME  
NUMBER OF SEQUENCES: 227  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/879,957  
FILING DATE: 13-Jun-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/630,915  
FILING DATE: 03-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie

REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-174  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 134:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 56 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 134:  
US-09-879-957-134

Query Match 4.9%; Score 309; DB 10; Length 56;  
Best Local Similarity 98.2%; Pred. No. 4.9e-09;  
Matches 55; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 907 EQLQALYPMRAKKDNLNFKSDVITVLEQDDMMWFGEVQGGKGFPPKSYVYKLI 962  
Db 1 EQLQALYPMRAKKDNLNFKNDVITVLEQDDMMWFGEVQGGKGFPPKSYVYKLI 56

Search completed: March 14, 2003, 12:17:46  
Job time: 479 secs



GenCore version 5.1.4 ps 4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2003, 11:05:51 / Search time 25 Seconds  
(without alignments)  
4664.442 Million cell updates/sec

Title: US-09-674-237A-3

Perfect score: 6269

Sequence: 1 MAQFTPTFGSGLDVMAITVE.....VGLFPSNVKLTITDMPDSQ 1213

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r73:\*  
2: p1r12:\*  
3: p1r3:\*  
4: p1r4:\*

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5063.5	80.8	1270	2 T09194	adaptor protein in
2	1735	27.7	1011	2 T13055	dynam associated
3	1698.5	27.1	1094	2 T13053	dynam associated
4	1333	21.3	1097	2 T31504	hypothetical prote
5	506.5	8.1	751	2 T34490	hypothetical prote
6	486	7.8	897	2 A54696	EGF receptor subst
7	485.5	7.7	896	2 S43074	epidermal growth f
8	410.5	6.5	1480	2 S48440	poly(AI)-specific r
9	403	6.4	1381	1 S45781	probable calcium-b
10	391.5	6.2	1407	1 S28589	trichovyalin - tab
11	383.5	6.1	1909	2 A45592	liver stage antige
12	381	6.1	1898	1 A45973	trichovyalin - hum
13	375.5	6.0	1549	1 A40691	trichovyalin - she
14	359.5	5.7	737	2 T15597	hypothetical prote
15	359.5	5.7	1017	2 T15598	hypothetical prote
16	342	5.5	2526	2 T20531	hypothetical prote
17	342	5.5	2722	2 T20532	hypothetical prote
18	340	5.4	2738	2 E88320	hypothetical prote
19	331.5	5.3	1069	2 T00377	protein F07A11.6 f
20	330	5.3	793	1 JH0628	KIA0642 protein -
21	327	5.2	1794	2 T38459	caldesmon - human
22	324.5	5.2	1181	2 C86349	hypothetical diver
23	323.5	5.2	408	2 T42650	F8K7.4 protein - A
24	321	5.1	665	2 JC7191	hypothetical prote
25	319.5	5.1	1027	2 T46481	85K c-Chl-interact
26	318	5.1	699	2 E84655	hypothetical prote
27	316.5	5.0	816	2 T17257	hypothetical prote
28	314.5	5.0	887	2 G88484	hypothetical prote
29	309	4.9	4574	2 G02520	protein F23F12.8 f
					plectin - human

30	308.5	4.9	1701	2 T09127	probable erythrocy
31	307.5	4.9	1110	2 I51116	NF-180 - sea lamp
32	306.5	4.9	771	1 A33430	h-caldesmon - chic
33	304.5	4.9	585	1 A24168	involucrin - human
34	304.5	4.9	958	2 S47179	hypothetical prote
35	302	4.8	1052	1 A44937	kinetoplast-associ
36	302	4.8	4684	2 A59404	plectin (imported)
37	299	4.8	1233	2 T30989	sarcone/threonine p
38	297	4.7	2101	2 A42184	nuclear mitotic ap
39	296.5	4.7	3488	1 T34418	hypothetical prote
40	296.5	4.7	4957	2 T03455	ALR protein - huma
41	296.5	4.7	5262	2 T03454	ALR protein - huma
42	296	4.7	4687	1 A39638	plectin - rat
43	295	4.7	729	2 T50989	hypothetical prote
44	294	4.7	746	2 T47237	myosin II heavy ch
45	293	4.7	2442	2 T08621	centrosome associa

## ALIGNMENTS

RESULT 1  
T09194  
adaptor protein interseclin - African clawed frog  
C.Species: Xenopus laevis (African clawed frog)  
C.Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000  
C.Accession: T09194  
R.Yamashita, M.; Hoffman, N.G.; Hardison, N.L.; McPherson, P.S.; Castragnoli, L.; Cesareni, U. Biol. Chem. 273, 31401-31407, 1998  
A.Title: Interseclin, a novel adaptor protein with two eps15 homology and five src homol  
A.Reference number: Z16605; MUID:99030416; PMID:9813051  
A.Accession: T09194  
A.Status: preliminary; translated from GB/EMBL/DBJ  
A.Molecule type: mRNA  
A.Residues: 1-1270 <YAM>  
A.Cross-references: EMBL:AF032118; NID:g2642624; PIDN:AACT3068.1; PID:g2642625  
A.Experimental source: cell type oocyte  
C.Function:  
A.Description: involved in endocytosis  
C.Keywords: endocytosis

Query Match	80.8%	Score 5063.5	DB 2	Length 1270
Best local similarity	80.6%	Pred. No. 2.6e-198		
Matches 980	Conservative 98	Mismatches 123	Indels 15	Gaps 8
QY	1	MAQFTPTFGSGLDVMAITVEERAKHDQFLSLKPIAGITTDQARNFFPQGLPQVLAQ	60	
DB	1	MAQFTPTFGSGLDVMAITVEERAKHDQFHLKPTAGITTDQARNFFPQGLPQVLAQ	60	
QY	61	IWALADNMNDGMDQVFSIMKLIKTLQGYQPLSTLPVPMKQPVVAISA--PARGIG	118	
DB	61	IWALADNMNDGMDQVFSIMKLIKTLQGYPLPSILPSMMLKQPVAMPAAVAVAGGMS	120	
QY	119	GIASMPPLTAAVAVPMGSIPIVVGMSPLVSSVPPAAVPLANGAPVYIOPLPAAFAHPAT	178	
DB	121	GIAGIPPLAAVAVPMGSIPIVVGMSPLVSSV--TVPLSNGAPAVIQSHAPAAH-SAT	177	
QY	179	WPKSSFSRSGSGSLNKLQKAGSFVVASAPPAEVAVPPSSSLKTRQLFNSHDKTMGS	238	
DB	178	LPKSSFSRVSAGSINIKLQKAGSFVPAAPLVVAEVAVPPSSSLKTRQLFNSQDKTMGS	237	
QY	239	HLTGQARTIMQSSGLPQALASIMNLSIDIDQDKLTAEFTILAMHLIDVMSQPLPPV	298	
DB	238	NITGQARTIMQSSGLPQALASIMNLSIDIDQDKLTAEFTILAMHLIDVMSQPLPPI	297	
QY	299	LPPEYIPPSFRRVSGSGSVISSSVVDQRLPEPSSSEDEQPEKKLPVTTEDEKRENF	358	
DB	298	LPPEYIPPSFRRVSGSGSLISSVSVVDQRLPEPSEDEQPEKKLPVTTEDEKRENF	357	
QY	359	RGSVLEKRRQALLQEQKKEORLAOLERAQEKREKROEQEKKROULEKQEKREL	418	
DB	358	RGSVLEKRRQALLQEQKKEORLAOLERAQEKREKROEQEKKROULEKQEKREL	417	

QY 419 EROEEERREKIEERREAKKELEROROLEWERNRROELLNORNKKEOGTVLAKRKLTLE 478  
 |||||  
 DB 418 EROEEERREKIEERREAKKELEROROLEWERNRROELLNORNKKEOGTVLAKRKLTLE 477  
 |||||  
 QY 479 FELBALNDKKHOLEGKLQDIRCRLATOROEIESTNKSREIRIAETIHLQOOLQESQOMLG 558  
 |||||  
 DB 478 FELBALNDKKHOLEGKLQDIRCRLATOROEIESTNKSREIRIAETIHLQOOLQESQOMLG 557  
 |||||  
 QY 539 RLIFREKQILSDQKOVONSLHRDILLTKRLAEKELAQOOREOLDEVEREPTSLQOE 598  
 |||||  
 DB 538 KMLPEKQSLDQKOVONSLHRDILLTKRLAEKELQOOREOLDEVEREPTSLQOE 597  
 |||||  
 QY 599 IDVENNOLKELREIHSKOOLQOKRSLERARLKOKERKSLERKEDAKORRVOERKO 658  
 |||||  
 DB 598 IDVENNOLKELREIHSKOOLQOKRSLERARLKOKERKSLERKEDAKORRVOERKO 657  
 |||||  
 QY 659 WLEHVQOEBOEPFRKPHBEDRLKREDSVRKKEAEERAKPEMOKQSLRPHPHOEPALAT 718  
 |||||  
 DB 658 WQDVKQOEE--RYKFODEEKEKEESVQKEVER--KPELOEKPNKPFHQPEPGKLG 713  
 |||||  
 QY 719 QAPSTTEKGLTISAQESVKVYVYRALYPPESRSHDEITIOPGDIYVDESQTEGPGWL 778  
 |||||  
 DB 714 QIPMNTTEKAPLTIN--QGDVAVVYVYRALYPPESRSHDEITIOPGDIYVDESQTEGPGWL 772  
 |||||  
 QY 779 GGBELKGTGMPFANYAEKIPENEVPTPAKVTDLTSAAPAKLARETPAPL--PVTSEPS 837  
 |||||  
 DB 773 GGBELKGTGMPFANYAEKIPENEVPTPAKVTDLTSAAPAKLARETPAPL--PVTSEPS 837  
 |||||  
 QY 838 TTPNNMADFSSTWPSSEKPEETNDMTMAOPSLTVPSAGOLQORSAFPAATGSSPS 897  
 |||||  
 DB 829 TNSNNMADFSSTWPSSEKPEETNDMTMAOPSLTVPSAGOLQORSAFPAATGSSPS 888  
 |||||  
 QY 898 PVLQGEKEVEGLQOALYPMRAKKNHNFNKSQDVTIVLEQODMMWFGEVQOGKMPFKS 957  
 |||||  
 DB 889 PVLQGEKEVEGLQOALYPMRAKKNHNFNKSQDVTIVLEQODMMWFGEVQOGKMPFKS 948  
 |||||  
 QY 958 YVKLISGVNRKSTIDTGPTESSPAKRVASPAKPAIPGEFTIAMTVSSBEGDILTFQ 1017  
 |||||  
 DB 949 YVKLISGVNRKSTIDTGPTESSPAKRVASPAKPAIPGEFTIAMTVSSBEGDILTFQ 1008  
 |||||  
 QY 1018 QGDVIVVTKKQDMMWTVGVGKSGVFPNNYVRLKDESGTAGTGLGKKPEIAQVIAS 1077  
 |||||  
 DB 1009 QGDVIVVTKKQDMMWTVGVGKSGVFPNNYVRLKDESGTAGTGLGKKPEIAQVIAS 1068  
 |||||  
 QY 1078 YAAATGPEQTLTAPGQLIIRKKNPGWMEGLQARKKRQIGMPFANYVLLSGTSTKIT 1137  
 |||||  
 DB 1069 YAAATGPEQTLTAPGQLIIRKKNPGWMEGLQARKKRQIGMPFANYVLLSGTSTKIT 1128  
 |||||  
 QY 1138 PTELKPTAVQAVQVIGMVDYTAQNDDELAFSGQIINVINKEDPMMWKEVSGVGLF 1197  
 |||||  
 DB 1129 PTEPKPTSLPPTCOVIGMVDYTAQNDDELAFSGQIINVINKEDPMMWKEVSGVGLF 1188  
 |||||  
 QY 1198 PSNVYKLTITDMDPSQO 1213  
 |||||  
 DB 1189 PSNVYKLTITDMDPSQO 1204  
 |||||  
 RESULT 2  
 T13055  
 dynamin associated protein isoform 160-2 - fruit fly (Drosophila melanogaster)  
 C:Species: Drosophila melanogaster  
 C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
 C:Accession: T13055  
 R:Roos, U.; Kelly, R.B.  
 J. Biol. Chem. 273, 19108-19119, 1998  
 A:Title: Dap160, a neural-specific Eps15 homology and multiple SH3 domain-containing pro  
 A:Reference number: Z17594; MUID:98334647; PMID:9668096  
 A:Accession: T13055  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1011 <ROO>  
 A:Cross-references: EMBL:AF054612; NID:G2996029; PID:G2996030; PIDN:AAC39139.1  
 C:Genetics:

A:Gene: Dap160  
 A:Cross-references: FlyBase:FBgn0023388  
 Query Match 27.7%; Score 1735; DB 2; Length 1011;  
 Best Local Similarity 33.4%; Pred. No 2.4e-63;  
 Matches 413; Conservative 202; Mismatches 332; Indels 288; Gaps 31;  
 QY 11 SLDWAITVEERAKHDOQFLSKPIAGFITGDAQNFPGSGILPOVLAQIWMALADMMND 70  
 |||||  
 DB 4 AYDAWAVTPRERLKYQEQFRALQFOAGFVTGAQAKGFLOQLPPLILGQIWMALADYDSD 63  
 |||||  
 QY 71 GRMDVEFSIMAKLIKLOGLQVLPSTLPYPMKQOPVAISAPAFGIGIASMP--LT 127  
 |||||  
 DB 64 GGMNINERISACKLINLKRMDVPEKVLPSL-----LSL-----TGDPSPMTPRGSTS 113  
 |||||  
 QY 128 AVAPVPMGSIPIVGMSPPLVSVPPAAVPPANGAPVYIOPLPAPFAHPAATWPKSSFSR 187  
 |||||  
 DB 114 SLSPID-----PLKGI--VAVAPVVPVAPVPA--VATVISP--PGVSVPSGPTPTSN--- 162  
 |||||  
 QY 188 SGPSQNLTKLQKQSFVVASAPPAEAVVPOSSRLKYROLFNSHDKTMGSLTGPQART 247  
 |||||  
 DB 163 --PSRHTSISERAPSISVN---QGEWAVQAQKRKYTVFANANDRTSGVLTGQARQ 217  
 |||||  
 QY 248 IIMQSLPOAQIASTIMNLSIDIQDGKLTAEFFILAMHLIDVAMSGOPLPVLPPEYIIPS 307  
 |||||  
 DB 218 VLVQSKLPQVTLAQIWTLSIDIDGRLNCDEFILAMPCEKAMAGEKIPTVLPQEWVPPN 277  
 |||||  
 QY 308 FRRVSGSGMSVYSSSVDOPLPEPSSSD-----EQPEKLP--VTFEDK 353  
 |||||  
 DB 278 LRKISRPO---SVSGVSRPSGQPSARSHASVSQSGVGVADDPAGLPQGTSPEDKR 333  
 |||||  
 QY 354 RKNFERGSELEKRGQALLLEQORKEERLAQLEAEORKEERKEOEQAKRQLEKOLE 413  
 |||||  
 DB 334 KENVYQGEHLDRRRKIMEDQORKEERKEERKEERKEERKALUEARKQOELEKOLE 393  
 |||||  
 QY 414 KORELEOROEERREKIEERREAKKELEROROLEWERNRROELLNORNKKEOGTVLAKR 473  
 |||||  
 DB 394 ROREIEMEKEBERKRELEAKKAEKELEROROLEWERNRROELLNORNKKEOGTVLAKR 453  
 |||||  
 QY 474 RYLEBELEALNDKKHOLEGKLQDIRCRLATOROEIESTNKSREIRIAETIHLQOOLQES 533  
 |||||  
 DB 454 NTQLNVELSTLNEKIKELISORICTRAGVTVKTVIDMRTQRTDSSEMSQLARIEQ 513  
 |||||  
 QY 534 QQMLGRILPEK-----QILSDQKOVONSLHRDILLTKRLAEKELARQOLR 582  
 |||||  
 DB 514 NAKLQILTERAKWAKKASKASALGENAQEO-----LMAFAHKOLLINQIK 563  
 |||||  
 QY 583 EQLDEVERETREKLOEIDVFNQOLKELE-----IHSKOOLQOKRSLERARLKOKERK 637  
 |||||  
 DB 564 DKVENISKEISKEDIMTNDVQMSLKAELALITCKEDLYKEYDVQRTSVLEIKYKRK 623  
 |||||  
 QY 638 SLELEKQKDAQRQVQERDKQVLEHVQOEEOQPRPKPHBEDRLKREDSVRKKEAEERAKP 697  
 |||||  
 DB 624 NETSVSSAWDT-----GSSSAW-----BETGTTVDP-----AVASNDISALAAP 664  
 |||||  
 QY 698 EMQDKQSRLLFHPQEPKATQAPWSTTEKGLTISAQESVKVYVYRALYPPESRSHDEI 757  
 |||||  
 DB 665 AYD-----LGGPAP-----EG-----FVKQVAVFERNANABEI 693  
 |||||  
 QY 758 TIOPGDIYVDESQTEGPGWLGGELKGTGMPFANYAEKIPENEVPTPAKVTDLTSAAP 817  
 |||||  
 DB 694 TFPVGDIIILVPLEONAEPLMGLGEINGHTGMPPEYVYKLEVEGEV-----A 739  
 |||||  
 QY 818 PKLALBETPAPLPVTSSESTTPNNWADPSSSTWPSSEKPEETNDMTMAOPSLTVPSA 877  
 |||||  
 DB 740 PAAV--EALVQADVAT-----YNDKIN-----SSIPAA 768  
 |||||  
 QY 878 GQLRQSAFTPTATGSSPSVPLQGEKEVEGLQOALYPMRAKKNHNFNKSQDVTIVLE 937  
 |||||  
 DB 769 S-----ADLTAA-----DV----- 778  
 |||||  
 QY 938 QODMMWFGEVQOGKGMFPKSYVKLISGVNRKSTIDTGPTESSPAKRVASPAKPAIPG 997  
 |||||

[illegible]

RESULT 3  
 T13053  
 dyamin associated protein isoform Dap160-1 - fruit fly (*Drosophila melanogaster*)  
 C:Species: *Drosophila melanogaster*  
 C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
 C:Accession: T13053  
 R:Roots, U.: Kelly, R.B.  
 U. Biol. Chem. 273, 19108-19119, 1998  
 A:Title: Dap160, a neural-specific Eps15 homology and multiple SH3 domain-containing protein  
 A:Reference number: 217594; MUID:98334647; PMID:9668096  
 A:Accession: T13053  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1094 <RCO>  
 A:Cross-references: EMBL:AF053957; NID:g2984714; PID:g2984715; PIDN:AAC39138.1  
 C:Genetics:  
 A:Gene: Dap160  
 A:Cross-references: FlyBase:FBgn0023388

Query Match	27.1%;	Score 1698.5;	DB 2;	Length 1094;
Best Local Similarity	31.6%;	Pred. No. 8.1e-62;		
Matches 417;	Conservative 201;	Mismatches 329;	Indels 371;	Gaps 33;

```

0Y 11 SLDVIAIVEEBEAKDQOFLSLKPIAGFTIGQANFFPOSQSLPOPVIAQIALADMND 70
Db 4 AYDMAVAIVREBELKXQOBFQALQPOAGFVTGAQAGFLLQSLPPLILQIITALDITSD 63
QY 71 GRMDVEFSIAMKLIKLIKQYOLBSTPLPVYKQOPVAISSAPAFGIGIASMP--LT 127
Db 64 GKMNINEFSIACKTLINLKLRGMDVPKVLPPSL-----LSSL-----TGDVPMTRGSGTS 113
QY 128 AVAPVPMGSIPIVYGMSPPVSVPPAAVPLNGAPVPIQPLPAFAHAPATMPKSSSR 187
Db 114 SUSPLD---PKGI--VPPVAPVVPVAVPVA--VATVISP--PGVSVBSGPPPPSN--- 162
QY 188 SGPGSLDNTKLQKQASFDVASAPPAEMAVPQSSRLKYQLFNSHDKMTSGHLTGPOART 247
Db 163 --PSPRHTSISRAPSTESVN--QGEMAVQAQKRYTVQVFNANDRTRSGVLTGSOARG 217
QY 248 ILMOSSLPQAQIASIWNISDIDQDKLTAEFTILMHLIDVMSGOPPLPVLPPEYIPPS 307
Db 218 VLVOSKLTPOVTLAQITWLTSDIDGDRLNCDEFILAMFLCEKMAAEKIPVTLPOEWVPN 277
QY 308 FRVRSGSGMSVISSSSVQORLPEPPSED-----EQOPEKPLP--VFEDFK 353
Db 278 LRKISRRE----SYSGVYSPRGSGOPASHASVSSOSGVGVVDADPTAGLPQOTSFEDKR 333
QY 334 RENFERGSVELEKRRQALLQQRKFEORLQJLERAEQERKEROROEAKROLLEKOLE 413
Db 334 KENYVAGQALDRRRKIMEDQORKEEREEREREREADKREKARLEAEKROOELEORLO 393
QY 414 KORELEORQEEERREKREIEREAKLELOROLWEERNRQELLNORNEGGSTVLKAR 473
Db 394 RORELEMEBEORKELEKAKRELEKOROOBEWQARIAENMAOKEEORVLTOKOH 453

```

Qy	474	RKTLPELELANKKQLOEGKLODIRCRLATROEIESFNKSRLEIRLEITHLOOOLES	533
Db	454	NTQJANVELSTJNKNIELBQRI CDTRAGVTNKTIVIDGRFRQDRTSMSEMSQLKARIKEQ	513
Qy	534	QOMLGRILPEK-----QIISDOLKOVOONSJLHRDSLITLTKRLLEAKELARQOR	582
Db	514	NAKLLOLTERAKWEAKSKASGALLAGENAQOEO-----LMAFAHKOLINDIK	563
Qy	583	EOLDEVERETRSLKLOEIDYFNNOJLKEJRE-----JHSKOOLOKORSLBAARLKQKEOR	637
Db	564	DKZENIKELIESKEDINTNDVQMSLTKELSLALLTTCKEDLYKEDVORTSVLEIKYRK	623
Qy	638	SLELEKQKEDAOBRVOERDKQWLEHVQOEOPRPKPHREEDRLKREDSVVRKEAEERAKP	697
Db	624	NETSVSAMD-----GSSGAW-----BETGTTVDPY-----AVASNDISALALAP	664
Qy	698	EMQDKOSRLPHHOEPAKLAQTQAPMSTTKGELTISAGESVVNYRYRALYPFESSSHEI	757
Db	665	AVD-----JGGPAP-----EG-----FVYQOAVYEFNANAEI	693
Qy	758	TIOPGDIWMTDESQTEGPBGMLGELKGTGWFPANPAEIKIPENEVPTPAKPVTLDTLSAPA	817
Db	694	TFVVGDIILVPLPEQNAEPCHLAGEINIGHTGWPEPESTVEKLEBGEV-----A	739
Qy	818	PKLARETPALPEVTSSEPTTPNNMADFSSTPSSSNEKEPBTNDWDTMAAOPSLTVPSA	877
Db	740	PVAHV--EAPVDAQVAD-----YNDNINT-----SSIPAA	768
Qy	878	QGLRQBSAFTPATATGSSSPBPVLGQEKVEGLQALYAPWRAKKNHLNFKNSDITVILE	937
Db	769	S-----ADLTAAG-----DV-----	778
Qy	938	QODMMWFGEVQOGKWFPPKSYVXLISGPVRKSTIDTGPTEPSALKEVAPPAKPAIG	997
Db	779	-----	778
Qy	998	EEFTAMTYESSQOGLTFOQGDVIVTKKDDMTGTIVGDKSGVFPSPNYVRKLD-----	1052
Db	779	EYVIAAPYESAEBGDLFSFAGEMVWVILKKEGBMWTTGISRTGWFPSNYYOKADVGTAS	838
Qy	1053	-----SEBGS-----	1053
Db	839	TAAAEVPESLDQETTLNGNAAVTAALVEAEOBOVYQPLPVQEBSEOPISSPGVGAEEAHED	898
Qy	1058	-----TAGKT-----GSLCKKPELIAQVIASVYATGPEOLT	1087
Db	899	LDTEVSQINTQSTQOSEPEPASYSRMSRSTSSMTPEMKRASEINQVIAAPYATSTEDLS	958
Qy	1088	LAPQOLILIRKKNPGWMBSELQARSKKQOIGFPPANNYKXLSPG--TSKITPYTELPTXA	1145
Db	959	LTRGOLIMIRKKDSDGMBEGELQAKRRROIGWFPATYKXVLQGRNSGRNTPVSGSRIE	1018
Qy	1146	VQPAVC--QVIGMTDYTAQNDDELAFSPKGOIINVUNKDEPDMKGEVSGOVGLFPNNY	1202
Db	1019	MTBQIDLKVALTPYKAQNDDELSPDKDILISVLGRDEPMMRGEINGLISGLFPNNY	1076

RESULT 4  
T31504  
hypothetical protein Y116A8C.36 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
R:Accession: T31504  
R:McMurray, A.  
submitted to the EMBL Data Library, October 1999  
A:Reference number: Z21041  
A:Accession: T31504  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1097 <WIL>  
A:Cross-references: EMBL:AL117204; PIDN:CBAS5138.1; CESP:Y116A8C.36  
A:Experimental source: clone Y116A8C





QY 416 RELEROREBERK-----ETERRAKRELEOROLEMERNRROELLNORNEQSTVV 469  
 Db 471 KOLEROGEATRRRLADYDTQIEQLESA---CKAQKETKEDEKRMQOIDEAKNAD--- 524  
 QY 470 LKARRKLEFELEALNDK--HOLEGLDODIRLATOROIESTNKSRELRJAETHIQ 527  
 Db 525 CKA-----NDEKEMELKEIEMLDNDFKTRVEIKETSORBQKABELT--- 569  
 QY 528 QOLQESQOMLGRLLPEKQILSDQLKQVONSILHRDLSLLTKRALKEKELARQOLR-EOLD 586  
 Db 570 -----TLERKE-ARDQIQHEKD 586  
 QY 587 -EYERERKSLQIEDVNNQKLEIREHISKOQLOKORSLAARLKQKEPKSLLEKOK 645  
 Db 587 AAIENTTKLEQVSDAVEKSEEMVQI-----LRSQORLSTVIDSLSDDDTVYGETAG 641  
 QY 646 EDQORAVQERDKQWLEHVQOEOPRPRKPHBEDRLKREDSVRKKKEERAKPEMQOKSR 705  
 Db 642 TSSQ-----NHVQPPDD-----FASARNPAD----- 665  
 QY 706 LPHHQBPAKLATQAPWSTTEKGPLTISAQESVYVYRALYPPESRSHDEITIQPDIV 765  
 Db 666 -----PFAQYDQ-----FGSSGH----- 678  
 QY 766 MWDESGTGEWGLGELKGTGMFPAN--YAEKI PENEVTPAKPVYDLSAPAPKLAIR 823  
 Db 679 -FPAAPFTDFPFAQGG-----FPDSGFAQASAPAK---PAP-----RPAPPKSAR 719  
 QY 824 ETPAPLPVTSS-EPTTPNNMADFS 847  
 Db 720 ETPVNDPFAPSQOSTOPAGFADFA 744

## RESULT 6

EGF receptor substrate eps15 - mouse  
 C,Species: Mus musculus (house mouse)  
 C,Date: 23-Mar-1995 #sequence\_revision 23-Mar-1995 #text\_change 05-Nov-1999  
 C,Accession: A54696  
 F,Razdoli, F.; Mutchiello, L.; Matoskova, B.; Wong, W.T.; Di Fiore, P.P.  
 Mol. Cell. Biol. 13, 5814-5828, 1993  
 A,Title: eps15, a novel tyrosine kinase substrate, exhibits transforming activity.  
 A,Reference number: A54696; MUID:93361014; PMID:7689153  
 A,Accession: A54696  
 A,Status: preliminary  
 A,Molecule type: mRNA  
 A,Residues: 1-897 <PAZ>  
 A,Cross-references: GB:L21768; NID:9404756; PIDN:AAA02912.1; PID:9404757  
 C,Superfamily: calmodulin repeat homology  
 C,Keywords: EF hand; phosphoprotein  
 F,48-80/Domain: calmodulin repeat homology <EF1>  
 F,160-192/Domain: calmodulin repeat homology <EF2>  
 F,223-255/Domain: calmodulin repeat homology <EF3>

## Query Match

Best Local Similarity 7.8%; Score 486; DB 2; Length 897;  
 Matches 205; Conservative 138; Mismatches 298; Indels 322; Gaps 31;

QY 15 WATTVBERAHDOQFLSLKRIAGFITGCDARNFQSGLPQVLAQITMALADNNNGRMD 74  
 Db 122 WAKSSEDKAKYDAIFPSLSFVDSGLSDKVKPVLLNSKLPVELIGVWELSDIDHDKLD 181  
 QY 75 QVEFSIAMKLKIKLQGYQVSTLPVPMKQPVAISSAPAFGIGIASMPPLTAVPVPM 134  
 Db 182 RDEFAVAMFLVGCAL-----KEPVIM 203  
 QY 135 GSIPVVMGSEPLVSVPPAAVPLANGAPVIOPLPAHPAATWPKSSSFSSGSGSQ 194  
 Db 204 -----SLPPALVPP-----SKR 215  
 QY 195 NTKLQKQKQSDVASAPPAEMAVPQSGRLKYROLFNSHDKTMSGHLTGQARTILMOSS 254  
 Db 216 KT-----WVSPAKAKYDEIFLTKDKMDGVSGLEVRRTFLTKGL 257

QY 255 POAQLASIMNLSIDIDQDKLTAAEFILAMHLIDVAM-SGQPLPVLPPEYIPSPFRVRS 313  
 Db 258 PSALLAHIMSLCDTGTGCKLSKQDQALAFHLINQKILKIGIDPPHSLTFEMIPPS---DRS 314  
 QY 314 GSGMVISSTSSVDOKLPPEPSSDEQOPKPLPTFPEKKRENEERSVLEKROALLE 373  
 Db 315 SLQKNITGSSPV-----ADFSAIK-----ELDTLNEIYD 344  
 QY 374 QORKEORLAQLEAEREREROREOEAKROLEKOLEKORELEROREBERREIERR 433  
 Db 345 LOREK-----NVEQDLKEKE-----DYVQRTSEVQ--DLQDEVORE 380  
 QY 434 EAKKELEROROLEMERNRROELLNORNEQSTVVLKARRKLEFELEALNDKQHEG 493  
 Db 381 SINLQKQAOQK-----QVQELLGE-----LDQKAOLEE 410  
 QY 494 KLQDIRCLATOROIESTNKSRELRJAETHLQOQLOESQOMLGRLLPEKQILSDQLK 553  
 Db 411 QLOEVRKKCAEBAOLISLK-----AEITSOESQISSYEBELK-----AREELSR 456  
 QY 554 VQONSILHRDLSLLTKRALKEKELARQOLREQDEVRE-----TRSKLOEIDVNNOLKE 608  
 Db 457 LQOETAQ-----LEESVSGKAQLEPLQOHLQBSQOELISSMQRLEKQDLETDNNQ--- 507  
 QY 609 LREHNSKOOLQK-----RSLAARLKQKEORKSLELEKOKEDAOORVQ 653  
 Db 508 -----SNWSSPQSVLVNGATDYCSLSTSSSETANFPMHAGQNNLSEPHQSSSVRS 562  
 QY 654 ER-----DKOWLEHVQOEOPRPRKPHBEDRLKREDSVRKKKA---EERAKPEMODK 702  
 Db 563 PEIAPSDVTDSEAVTVAGNEKVTPR--FDQDKHKSKEDEPENVSSSLTDAVADTNLDF 620  
 QY 703 GSRLF-----HPHOEPAKLATQAPWSTTEKGPLTISAQESVYVY 743  
 Db 621 QSDPVGSDPFKDPFGKIDPFQGDPPFGSDPPFASDCEFKOTSTDPTTSTDPFASSN 680  
 QY 744 RALYPPESRSHDEITIQPDIVMWDESGTGP--GWLGLGELKGTGMFPANVYAEKIPENE 801  
 Db 681 SSNTSVETWIKNDPFAPOGTVAASAASDADTPFASVFGNESFG--DGFADFSTLTKANNED 739  
 QY 802 VPTP-----AKVYDLSA-----PA--PKIALRETPAPLPVTSSESTTPNN 842  
 Db 740 AFNPITSSSTSSVTIAKMLEETASKSSEDVPAALPVKGPVTPRCPPP-----GKKEPIN 794  
 QY 843 WADFSST-----WPSSNKEPETDWMWDTMAAPSLTVPSAGQLRQSAFTPATGSS 895  
 Db 795 KLDSSDPLKLDNPPQPPGNDSPKEXPDWFC-----DPTSSSTTNKE 838  
 QY 896 PSP 898  
 Db 839 ADP 841

## RESULT 7

S43074  
 epidermal growth factor receptor substrate - human  
 C,Species: Homo sapiens (man)  
 C,Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 01-Dec-2000  
 C,Accession: S43074; I38525  
 R,Bernhard, O.A.; Mauchauffe, M.; Mecucci, C.; van den Berghe, H.; Berger, R.  
 Oncogene 9, 1039-1045, 1994  
 A,Title: A novel gene, AF-1p, fused to HRX in t(1;11)(p32;q23), is not related to AF-4/  
 A,Reference number: S43074; MUID:94181254; PMID:8134107  
 A,Accession: S43074  
 A,Status: preliminary  
 A,Molecule type: mRNA  
 A,Residues: 1-896 <BER>  
 A,Cross-references: EMBL:Z29064; NID:9470034; PIDN:CAA82305.1; PID:9470035  
 R,Wong, W.T.; Kraus, M.H.; Carlomagno, F.; Zelano, A.; Druck, T.; Croce, C.M.; Huebner, J.  
 Oncogene 9, 1591-1597, 1994  
 A,Title: The human eps15 gene, encoding a tyrosine kinase substrate, is conserved in evo]



```

QY 161 GAPPVIO-----LPAPAPATWPKSSFSRSGPSQL-----194
D 506 GA---IOSYTGSGVMPQOSGPASW--PMLSFNOGLOSQLTGLQOPQTFGLPSPNFS 560
QY 195 -----NTKLQKASFPVASAPAPAAEMAVDQSSRLKYROLFNSHDMTMSGHLT 241
D 561 ATMPETAKTGFGNNELYTTSNFNNLIDNSSQDKISTEKSILFYKIFETDTQNGKGLD 620
QY 242 GPQARTILMOSSLPQOALASIMWLSIDODGKLTAEFFILAMHLIDVAMSGQPLPVLPP 301
D 621 SPFAVEIFRKSGLNRADLEQIMWLCIDINTGQLNKQEFALGMHLVYGLKNGKIPVLLS 680
QY 302 EYIP-----PSF-----RRVSG--314
D 681 SLIPSTKLIDNLKQLTKEPTTKEKPSFGKIDALSYYKNNDDVLPNYRNRRKVVSAKN 740
QY 315 ---SGMSVSSSSVDQRLPEPSSDEQPEKKLPYTFEDK---KRENERSVLEKR 367
D 741 EEOSSFPSPAKSVNH-----SSSTLQTDIDISVDKIVEKTKAPKYAGFSR---EINK 791
QY 368 RQALLLEQORKEOR-----LAQLERAEOERKEREQOEAKROLE 407
D 792 NIASLEMEIKNINPENCYDSSIPDLTSRFDALIAKLPLFNEISTIDNEITNAKIQLY 851
QY 408 LEQOLEKQRELEQREBERKEIERRBAARELQROLEWERNRROELINORKEQEGT 467
D 852 RKR---NPSSIIISGPGEITENDRKAKSRALLR-----ARMSALTGKSTESDS 899
QY 468 VVAKARKTLEFE--LEALNDKKHQLESGKODICRLATQROELIESTNKRRLIAITIL 526
D 900 LSWEDQOQSAEIRIQOENKQOEI---IKDIRSSIDISASLKTSTGSM---ISN- 951
QY 527 QOQLOESQMLGRILPEKQILSDQLKVOQNSLHDS-----563
D 952 -QFERMEFEGTIGLEDGREFL--DLKSNKSKVTESSPFVPSSTPPVDRSSPSYSGP 1009
QY 564 LTLTKALAEKELARQOLREQLDEVERETRSKLQEIIVEN--NQLKELEIHSKQOL-- 618
D 1010 KTAEBRAAYLKEQAKKMKKELAKPKNRNRYVQSSRSISSENSRQPOQIAGSSMLVBP 1069
QY 619 -----OKORELEAARLKQKEKESLLEKQKE-----646
D 1070 RATPFOBEKVEVAQPTQPVQSTQPVQPTQPVQPTQPVQPTQPVQPTQPVQPVQNY 1129
QY 647 DAQRVQERDKQWLEHVQOEQPRPRKHEDERLKR-----EDSVKKEAE--693
D 1130 NAKQESDDEDEDEBEKRLQELKRLKKADEKELALARKQIEDAQNESDEEFNGKD 1189
QY 694 -----RAKP-----EMQDKQSLFHPHQBPA--KLATQAPWSTTEKGPLTISA 734
D 1190 NFGHVAVVPOAAVAPASAFQSNSTNAPRSVHAAYTPAAGKNTGLPSTTMGNHPYFKDA 1249
QY 735 QESVKKVYVYALYPFSRSHD--EITQPGDIIWVDESQTEPWLGLGELKGTGMFPANV 793
D 1250 SAS-----STSTPFAAEMQRRIQG---LDDEB--DDMSDEDESDNNR---VAV 1292
QY 794 AEKIPENEVPTP---AKPYTDLTSAAPKALARETAP--LPVTSSEPTPTNNWADS 847
D 1293 DNKVEEAKIGHPHARAPV---TAAPLPBV---TPVPAVVPVPAINTSNESSSIPPIA 1345
QY 848 STWPSSSNE-----KPEITMDTMAAQSPLTVSAQLRQRS--ATPATATGSSPSPLV 901
D 1346 PIPPSVTOEPVPLAPPLPAVDGFOEPPIPSADAIATAVGKSSSPALAGVLPFP--- 1402
QY 902 QGKVEGLQALYPMRAKDNHLNFK-----SDVITVLEQDQMMWFGEVQCK 951
D 1403 --PPLPTQASTSEPIIAHDVNYNGAEKGTGAYGSDSDVDVLSIPES-----VGTDEEBE 1455
QY 952 GWFPSYVXLISGP 965
D 1456 GAQPVSTAGIPSP 1469

```

## RESULT 9

S45781

probable calcium-binding protein YBL047c - yeast (*Saccharomyces cerevisiae*)

N.Alternate names: hypothetical protein YBL0520

C.Species: *Saccharomyces cerevisiae*

C.Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 19-Apr-2002

C.Accession: S45781; S50284; S45782; S39841; S37339; S42498

R.Goifeau, A.; Jomiaux, J.L.; Purrelle, B.; Skala, J.; de Wergifosse, P.; van Dyck, L.

submitted to the Protein Sequence Database, August 1994

A.Reference number: S45745

A.Accession: S45781

A.Molecule type: DNA

A.Residues: 1-961 <GOF>

A.Cross-references: EMBL:Z35808; GSPDB:GN00002; MIPS:YBL047c

A.Experimental source: strain S288C

R.de Wergifosse, P.; Jacques, B.; Jomiaux, J.L.; Purrelle, B.; Skala, J.; Goifeau, A.

Yeast 10, 1489-1496, 1994

A.Title: The sequence of a 22.4 kb DNA fragment from the left arm of yeast chromosome II

NA-binding protein.

A.Reference number: S50284; MUID:95176707; PMID:7871888

A.Accession: S50284

A.Molecule type: DNA

A.Residues: 1-961 <DEF>

A.Cross-references: EMBL:X78214

A.Experimental source: strain S288C

R.Dubois, E.; El Bakoury, M.; Glansdorff, N.; Messenguy, F.; Pierard, A.; Scherens, B.;

submitted to the Protein Sequence Database, August 1994

A.Reference number: S45782

A.Accession: S45782

A.Molecule type: DNA

A.Residues: 579-1381 <DUB>

A.Cross-references: EMBL:Z35808; GSPDB:GN00002; MIPS:YBL047c

A.Experimental source: strain S288C

R.Scherens, B.; el Bakoury, M.; Vierendeels, F.; Dubois, E.; Messenguy, F.

Yeast 9, 1355-1371, 1993

A.Title: Sequencing and functional analysis of a 32 560 bp segment on the left arm of ye

A.Reference number: S39824; MUID:94205266; PMID:8154187

A.Accession: S39841

A.Molecule type: DNA

A.Residues: 579-1381 <SCH>

A.Cross-references: EMBL:Z23261; NID:G313733; PIDN:CAA80797.1; PID:G313748

A.Experimental source: strain S288C

C.Genetics:

A.Gene: SCD:EDL1; MIPS:YBL047c

A.Cross-references: SGD:S0000143

A.Map position: 2L

C.Superfamily: yeast probable calcium-binding protein YBL047c; calmodulin repeat homology

C.Keywords: calcium binding; EF hand; transmembrane protein

F.167-199/Domains: calmodulin repeat homology <EFL>

F.166-576/Domains: transmembrane #status predicted <TMM>

## Query Match

Best local Similarity 6.4%; Score 403; DB 1; Length 1381;

Matches 286; Conservative 199; Mismatches 486; Indels 458; Gaps 57;

```

QY 3 QPPTP-OGS-----LDVVAITVEBRAXKHDOFLSLKPIAGITGDQARNFFPQSLP 54
D 109 QNPAMQSGSATGNTNNTDIPALSSNDIAKFSQLPDRTAGVQVADADKXDFLKLRLP 168
QY 55 QPVAIQIALADNMNDGMDQVEFSIAMKLIKLGQYQYLPSTLPVYKQOQVPAISSAPA 114
D 169 NQTLGEIHALCDRDASGVLDKSEFIMAMVYLQLCMSHHPNMTTPVAVLPQ----- 219
QY 115 FGIGIASMPPLTAVAPVMSGSPVYVGVSSVPPVAVPLANAPVYIOPLPAFAH 174
D 220 -----LMDSI---RLPEVYVAVQPRRTPLSANGT 245
QY 175 PAATWPKSSFSRSGPSQLNTKLQKASFPVASAPAPAAEMAVDQSSRLKYROLFNSHDK 234
D 246 GVSLSLTHSTISRLSTGAFSA-----ASDMSLSFEKKQOQFDALFDSLDK 290
QY 235 TMSGHLTGPQARTILMOSSLPQOALASIMWLSIDODGKLTAEFFILAMHLIDVAMSGOP 294

```

```

Db      291  CHAGSLSSAVLPFPFSSRLNQLATIMDLADHNNAEFTKLEFALIMLLIOKKNAGVE 350
QY      295  LPLPLPE-----YIPSPFRVRSGSGMSVSSSS-----324
Db      351  LPDVIENELLQSPALGLYPNPPLPQOQSAFOIALPSRASKPSLQDMPHOVSAVANVQPI 410
QY      325  VDQRLPE-----BPS-----SEDEQOP 341
Db      411  VPQVLPNQSNNGSLNDLLALNPFSFSPPTKACTVQNTNNSFYDNNNGAQLTQQOOP 470
QY      342  EKKLPVTFPEKKRNFERSVELEKROALLQOORKEOEL-----382
Db      471  QQPPLTHSSSGLKKTPTPS---NFGQSIIEKEPEEQOLRESSDTSAPPPPKHAS 526
QY      383  -----382
Db      527  SPVKTASTTLTPQVPNFSVSPMPAGATSAATGAAGAAGAAAGASAFSSSNNAFKN 586
QY      383  -----AQLERAEQERERERQEOAKRQOLEKQOLEKQERELERQEEERKEIER 432
Db      587  QDLFADGEASAKLSNATTEMANLSQVNSLSKQASITNDKKSRAQOLKRVTEKKSIOI 646
QY      433  REAKRELERQOLEMERNRROELLNQKQOEGTVLKKARKTLEPELEALNDKKHOLE 492
Db      647  KLNLRKSTHDQNVKOTOLEAQVL--QVKNENE-----TLAQOLVSEANYHAIE 694
QY      493  GKLODIRCLATROEIESTNKSRELIAEI---THLQOOLQSOQMLGRLLPEKOILS 548
Db      695  SKINE-----LTTDLQESQTKVAELKEQITLNSWTASLOSQLNKEQ-----737
QY      549  DQKQVQONSJHRDSLTLTKALFAKELARQOLEQDEVERETRSKLOEIDVPNNOLKE 608
Db      738  -QVQKQ-----ERSMVDNSKQLELNOVTYANLOKEIDGQ-----EKISYLLTKQKE 783
QY      609  LREHSKQOOLQOKRSLLEAKRLKQKEQ--RKSLLEL---EKQKEQACRQVQERDKOWLEHV 663
Db      784  LND-----YKTYEBQAOLOAKYQDLSNKDITDLTRBKQLEBRNNOIQIEQENLHQHV 837
QY      664  QOEQOPRRKHEEDRLKREDSVRKKEAEERAK-----PEMODYOSKLFPHQEP-- 713
Db      838  SKLOE-----MPDLSQKASFEKADQELKERNIEYANNVRELSEKQML--AMGQLPED 890
QY      714  -----AKLATQAPWST---TEKGPL---TIS-----AQESKVVYVYALVPFERSHDEI 757
Db      891  AKOIIASASNTDTTTEKATSRGVNHEVTYSKFVETTVENSNNLVNVRVKODEETERTES 950
QY      758  TIQPGDI--VWVDESQTEPFWLGGELKQKQGMFPANYAEKIPE-----NEVPTP--AKPV 809
Db      951  DVFPRDVTLLSQSDSEVANNTNGQSGNETANP--NILETILSDRFQDGLNEYGIPRSQSL 1009
QY      810  TDLISAPKALARETPAPLEVTSEPEPTTPNNWADFSSTPSSNKEPFDNDMDTMAAQ 869
Db      1010  TSSVANNAPQ--SVRD--DVELPELTLEERDTI--NNTANDNT---GNLSHPGEAEATPAT 1062
QY      870  PSLVTPSAGQURQSAFTPATATGSSPEPVLG--QGEKVEGLQAO--ALVPMRAKQDN- 923
Db      1063  ASTDVL-----NETTEVLEDSGTTGRANSNEDGEVSSIQSPKISAPQ--KAKTINE 1114
QY      924  -----HUNFKSDVITVLEQODMMWGEVQOGKQWPFKSYVKLISGP-----965
Db      1115  EFPPIQELHIDESSSSDDE---FEDTRE---IPSAIVKTLQTPYNAOPTSSLEIH 1166
QY      966  ---VRKSTSIDQTPESASIKRVASPAKAPAI--GEPIFAMTYTSSSEQGLTFQOGD 1020
Db      1167  TEQVIKTPAPGTSBHNHGNKSKASTNSILPVKQEFDDFAGL-----EQAAVEEDNG- 1219
QY      1021  VIIVTKKDDMMWTGVDKSGVFPSPNYVRLKDS-----GSGTAGKTGSLGKKEPEI 1071
Db      1220  -----ADSESEFENVANAGSEQETIDHKDLDDELQMANAFTGLTSSSNPTI--PKQV 1272
QY      1072  AQVIASVATGPBQULTLAPGQILIRKKNPGGWEGELQAGKKRQIGMFPANYVKLSP 1131

```

```

Db      1273  QQ-----OSTSDPAQV-----SNDEWD-EIFA-----GF-----1295
QY      1132  GTSKITPTLEL---PKTAVQPAVCQIGWYDTAOND--ELAFSKQIIN 1176
Db      1296  GNSKAEPTKVAFTPSIQOP-----IPLKNDPVIDASLSKGPVIV 1334

RESULT 10
S28589
trichohyalin - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 12-Mar-1993 #sequence_revision 01-Mar-1996 #text_change 22-Jun-1999
C:Accession: S28589
R:Rietz, M.J.; Rogers, G.E.
submitted to the EMBL Data Library, December 1992
A:Description: Examination of the gene encoding rabbit trichohyalin.
A:Reference number: S28589
A:Accession: S28589
A:Molecule type: DNA
A:Residues: 1-1407 <PIE>
A:Cross-references: EMBL:Z19092; NID:91746; PIDN:CAA79519.1; PID:91747
C:Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root she
Covalent modifications to this protein include conversion of arginine to citrulline and t
C:Genetics:
A:Introns: 46/3
C:Superfamily: trichohyalin; calmodulin repeat homology
C:Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat
F:49-81/Domain: calmodulin repeat homology <BP2>

Query Match      6.2%; Score 391.5; DB 1; Length 1407;
Best Local Similarity 29.7%; Pred. No. 1.2e-08;
Matches 132; Conservative 97; Mismatches 140; Indels 75; Gaps 18;

QY      326  DQRLPEPSSSEDEQPEKKLVPTEDKRNENFRGSVLEKRRQALLQO-----375
Db      198  EEFIEEQLRRRQQLKRELREEQQRERREQHERALQEEBQLRQRWREPREBQ 257
QY      376  -----RKEORLQLERAQERKERERQEOEAKQOLELEKQERELERQEEER 426
Db      258  QLRLELEIRERQRLRQLEERREQQLRQELRQLEERREQQLRLELEIRERQRLQEE 317
QY      427  RKE-----IERRE-AAKREL---ERQOLEMERNRROELNQRNKEQ--EGVILKAR 473
Db      318  RREQLQEEERREQQLKRELLEIRERREQLQEE--ERRQLAEVRERARRGSLRRW 376
QY      474  RKTLEPELEALNDKKHOLEGKLODIRCLATQROEIESTNKSRELRIAEITHLQOOLQ-- 531
Db      377  QQLSEAGARQSKVSPRRQEEQSLFQDOERKQ---RQERRELEQARRQOQOAE 432
QY      532  -ESQMLGRL-----IPEKQILSDQKQVQ-----NSLRDSSLTLKRALEAKELARQ 580
Db      433  EESERRQQLSARPSLRRLQRLAEERQEOQRFREERQERREERQQLFLLEBQOLQRE 492
QY      581  LREQLDE-----VERETSKQOEL---DVFNQOLKE--LREIHS-----KQOLQKRS 623
Db      493  RAQQLQEEBDSFQEDERRRRQEOQRPQGTWQLOEABQRHHTLYAKPGQOEQLREEB 552
QY      624  LFAARLKQEKQERKSLLELEK--QKEDARQVQERDKQW--LEHVQOEOPRRKPHEDR 679
Db      553  LQREKRRQ--ERRREYREBEKQLQREDEKRRQERROVRQRLBELRLQEBQDLRLKREBQ 611
QY      680  L--KREDSVRKKEAEERAKPEMQ 700
Db      612  LIQEREERLRRQERERKLRREBQ 635

RESULT 11
A45592
liver stage antigen LSA-1 - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 22-Nov-1993 #sequence_revision 02-Dec-1994 #text_change 09-Jun-2000
C:Accession: S24597; A45592; S24592; S34842; B45592; C45592; D45592
R:Zhu, J.; Hollingdale, M.

```



Db 804 ARPLERQREKQLRAEERQREQ 826

## RESULT 13

A40691

trichohyalin - sheep

C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)

C&gt;Date: 21-Sep-1993 #sequence\_revision 01-Mar-1996 #text\_change 22-Jun-1999

C/Accession: A40691, A34209, S32633

J.Fietz, M.J.; McLaughlan, C.U.; Campbell, M.T.; Rogers, G.E.

J. Cell Biol. 121, 855-865, 1993

A&gt;Title: Analysis of the sheep trichohyalin gene: potential structural and calcium-binding

A/Reference number: A40691, MUID:93260018, PMID:7684041

A/Accession: A40691

A/Molecule type: DNA

A/Residues: 1-1549 &lt;FIE&gt;

A/Cross-references: EMBL:Z18361, NID:g295940, PIDN:CAA79165.1, PID:g295941

A/Note: sequence extracted from NCBI backbone (NCBI:P132511)

J.Fietz, M.J.; Prestland, R.B.; Rogers, G.E.

J. Cell Biol. 110, 427-436, 1990

A&gt;Title: The cDNA-decoded amino acid sequence for trichohyalin, a differentiation marker

A/Reference number: A34209, MUID:90130632, PMID:2298812

A/Accession: A34209

A/Molecule type: mRNA

A/Residues: 1016-1151, 1205-1257, 1281-1398, G', 1400-1549 &lt;F12&gt;

A/Cross-references: GB:X51695, PIDN:CAA5992.1, PID:g1828

A/Comment: Trichohyalin is a protein of the medulla of the hair and of the inner root she

Covariant modifications to this protein include conversion of arginine to citrulline and

C/Genetics:

A/Intons: 46/3

A/Note: single copy gene

C/Superfamily: trichohyalin; calmodulin repeat homology

C/Keywords: calcium binding; citrulline; EF hand; hair; tandem repeat

F/49-81/Domain: calmodulin repeat homology &lt;EF2&gt;

F/387-851/Region: 28-residue repeats

F/886-1519/Region: 23-residue repeats

Query Match

Best Local Similarity 27.6%; Pred. No. 5.9e-08;

Matches 116; Conservative 101; Mismatches 114; Indels 89; Gaps 16;

332 EPSEDEQPPKTLPTPE-----DKRPNFSGVLEKRRQALTEQQRKOEHLAQLR 387

Db 383 KPAGREYREBEOLRLKEKQLQREKRRQREKQYREVLQRE--ERLQREBEOLQREER 440

388 AEOERKERERQ-----DOAKRQLELEKO-----LEKORELEROREER 426

Db 441 EKRRRQREKQYLEKVELMEBEOLQREBERKRRQREKQYLEKVELMEBEOLQREER 500

427 RKELERRAAKRELERQOLEMERNRROELLNORNKQOEGVULKARKTLEFELALND 486

Db 501 ROEBERQYLEVELQOEOLQREERK-----RRQRE-----ROYLE----- 538

487 KKHOLEGLQDIRLATOROEISTNKSRLIAETIHLQ--OOLQSOOMLGRLLPEK 544

Db 539 -KVELQOEQ-----LQROREKRRQREKQYLEKVELQOEOLQROERQKRRQREK 590

545 QILSDOLKVOONSUHRDSSLTLKRALFAKELARQOLREOLDEVERETRSKLOEIVFNN 604

Db 591 QYL-EKVELQOEOLQROERERKRRQREKQYLEVELQOE--EYVOROREK-----RR 641

605 QKLELREIHSKOOLQOKSLAARLKQOEKRSLELEK-----OKEDAQOR 650

Db 642 QEREROYLEKLOREBEOLQOEOLRERERKRRQREKQYLEVELQOEOLQROERERK 701

651 RVQERDKQWL--EHVQOEOPRPKPHEDRLKREDSVKKKAER--AKEMQDKQSL 706

Db 702 RROEREROYLEKELOREOE--RLQOREKQLOREDEKRRQVVERKYVEELQOEEDRL 758

## RESULT 14

T15597

hypothetical protein C25A11.4b - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C/Accession: T15597

R.Favell, T.

submitted to the EMBL Data Library, October 1995

A/Description: The sequence of C. elegans cosmid C25A11.

A/Reference number: Z18375

A/Accession: T15597

A&gt;Status: preliminary; translated from GE/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-737 &lt;FAV&gt;

A/Cross-references: EMBL:U39650, NID:g1049376, PID:g1049381, PIDN:AAA80392.1; CESP:C25A11

A/Experimental source: strain Bristol N2

C/Genetics:

A/Intons: 65/3; 86/3; 126/3; 433/2; 494/2; 711/2

Query Match

Best Local Similarity 23.6%; Pred. No. 1.1e-07;

Matches 179; Conservative 142; Mismatches 270; Indels 169; Gaps 32;

236 MSG--HLNPGQARTILMS--SLPQALASIMNLSIDIDQDKLTAEFFILAMHLIDVAMS 291

Db 7 QSISSREDLSEHATSSSTVREIPVHRAPSTAP--SHSSVFEYHMMPTTSTYHHVET 63

181 KSSSSRSGSGQNTKQKQSPVYASPPAAEMAVQSSRLKYROLFNS-----HDKT 235

Db 7 QSISSREDLSEHATSSSTVREIPVHRAPSTAP--SHSSVFEYHMMPTTSTYHHVET 63

236 MSG--HLNPGQARTILMS--SLPQALASIMNLSIDIDQDKLTAEFFILAMHLIDVAMS 291

Db 64 PDEYVRRHVMRTITSTALSTQPL----- 91

292 GQPLPVLPEYIP-PSRRVRSGSGMSVISSSYDQPL-----PEESEDQ--Q 339

Db 92 GPAPPL--DRLLPYPTTTTGGGRT-REKTVYVYVYHDIIEGERRIRDQARQ 148

340 QPEKTLPTVEFKKRNFGSVLEKRRQ-----ALLQORKEORLALQLEAEOR 392

Db 149 QEOQRDRREDNARILIAQREHOMERLREOONLSERLALREKADKRLQ--RLLRQ 207

393 KEREROQ-----EAKROLEKOLEKORELEKOREERKEIERREAAKRELER 442

Db 208 REKKREEDRLSEIIRLAEFEALRRALREKER-IDEKAEKEEKTKMERLERERRARLER 266

443 QQLMEWRNRQ---ELLNORNKQOEGTV---VLKARKTLEFELALNDKXKQLEK 495

Db 267 ER-LEBERQKAKATTEIERRRHEHIEIERIRKREIEBERRREKKAED--- 322

496 QDIRCLATOROEISTNKS--RELRIAEITHLOOLQSOOMLGRLLPEKQILSDOLKOV 554

Db 323 ---RLRLERLEIRIERRELELAREFOLELQREAEADR-----EROLDEAREM 371

555 -QONSUHRDSSLTLKRALFAKELARQOLREOLDEVERETRSKLOEIVFNNQ--KELREI 612

Db 372 RREBERRAELVADVHROAERERLRRQREBERERLRLQOQKIDMERIDIAERER 431

613 HSKOOLQOKSL--FAARLKQOEKRSK---ELEKQEDAAORVQERDKQWLEHVQOEQ 668

Db 432 EKKEBERREFELIEMARKKEARDRLDEMEREVREBERREKRRQREORRIAAEKER 491

669 PRPKPHEE-----DLKREDSVKKKAERPAKEMQDKQSL 706

Db 492 KROREBEETIARINELORAAAARQORNAELDRORDELORK--AOELSEREMEKERD 550

707 PPHQEPALATQAWSTTEKGPLTISAQESKVYVYALPPFSRSHDETITQGDIVM 766

Db 551 RERANBAOLADLLERHNOILIRENERREAVERRANRL--DRRSQDL-----DHIV 603

767 VDESQTGEBGWLGGELKKGTFPANYAEKIPENEVPTPAKVTDLTSAPAKLARETP 826

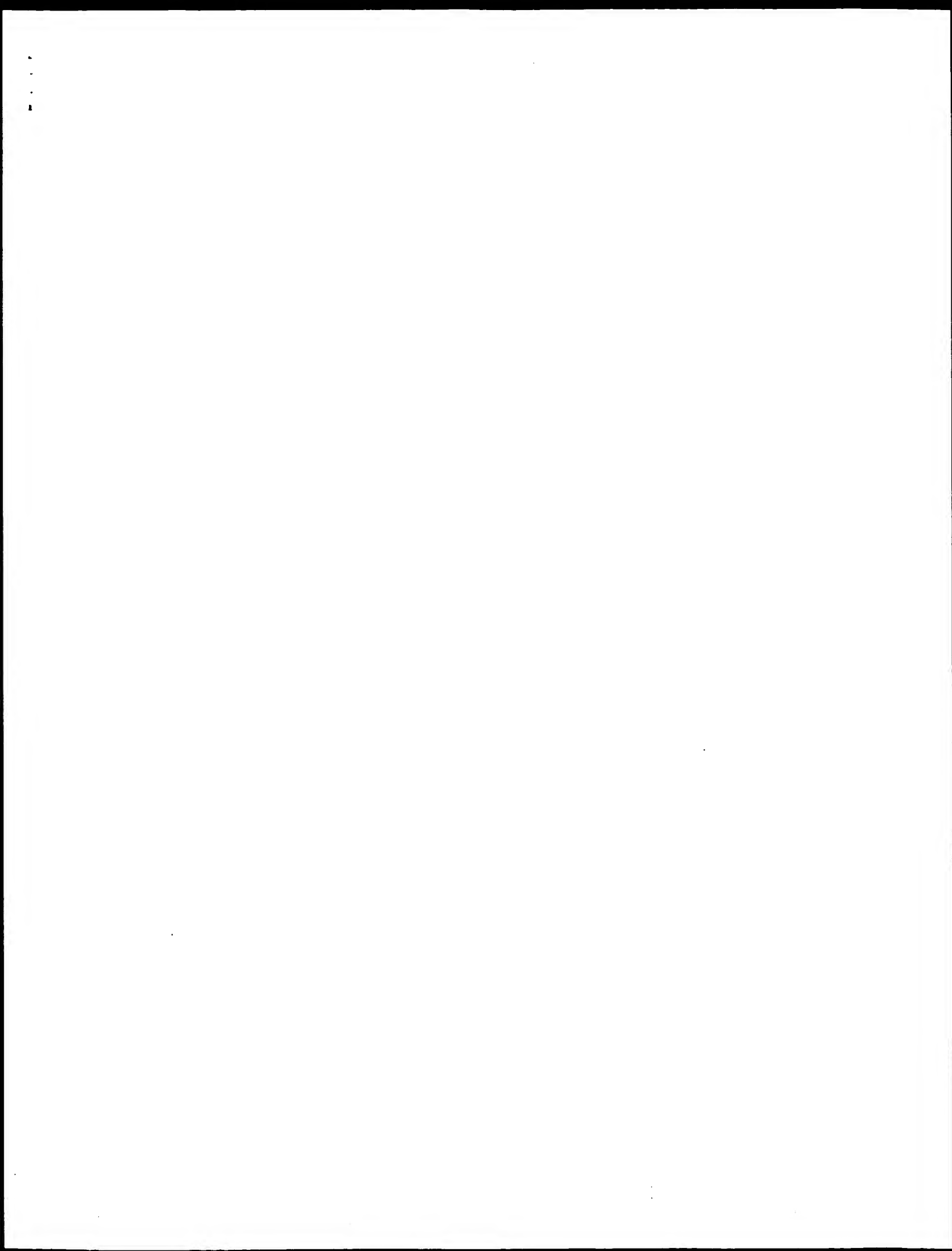
Db 604 RERSEKQ-----FELEKRRLLAEKAMARKK--NHLTSS-----TLAKLT 643

827 APLPVTSSEPTTP-----NNWADPSSTWPESS 854

Db 644 QPMYYTTRPEVTTKVEROVIERIDRNWVEDVYAPSO 683







GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2003, 08:18:17 ; Search time 38 Seconds

(without alignments)  
1323.968 Million cell updates/sec

Title: US-09-674-237A-3

Perfect score: 6269  
Sequence: 1 MAQFPFGSLDVMATVE.....VGLFPSNYVKLTMDMPSEQ 1213

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6269	100.0	1714	1 ITN1_MOUSE	Q920R4 mus musculu
2	6021	96.0	1217	1 ITN1_RAT	Q9wwe9 rattus norv
3	5842.5	93.2	1721	1 ITN1_HUMAN	Q15811 homo sapien
4	5063.5	80.8	1270	1 ITN1_XENLA	Q42287 xenopus lae
5	3068	48.9	1696	1 ITN2_HUMAN	Q92mz3 mus musculu
6	3049.5	48.6	1658	1 ITN2_MOUSE	Q920R6 mus musculu
7	486.5	7.8	896	1 EP15_HUMAN	P42866 homo sapien
8	466	7.8	897	1 EP15_MOUSE	P42867 mus musculu
9	410.5	6.5	1480	1 PAN1_YEAST	P32521 saccharomyc
10	403	6.4	1381	1 YBET_YEAST	P34216 saccharomyc
11	391.5	6.2	1407	1 TRHY_RABIT	P37709 oryctolagus
12	381	6.1	1988	1 TRHY_HUMAN	Q07283 homo sapien
13	375.5	6.0	1549	1 TRHY_SHEEP	P22793 ovine aries
14	330	5.3	793	1 CALD_HUMAN	Q05682 mus sapien
15	327	5.2	1794	1 YAVI_SCHPO	Q10172 schizosach
16	314.5	5.0	887	1 YLX8_CAEEL	P46504 caenorhabd
17	306.5	4.9	771	1 CALD_CHICK	P12957 gallus gall
18	304.5	4.9	585	1 INVO_HUMAN	P07876 homo sapien
19	304.5	4.9	958	1 IF3A_TOBAC	Q40554 nicotiana t
20	302	4.8	4684	1 PLEI_HUMAN	Q15149 homo sapien
21	298	4.8	4473	1 PLEI_CRIGR	Q91555 cricetulus
22	296	4.7	4687	1 PLEI_RAT	P30427 rattus norv
23	283	4.5	733	1 VINE_MOUSE	Q91128 mus musculu
24	283	4.5	1805	1 NEST_RAT	P21263 rattus norv
25	281.5	4.5	544	1 INVO_AOTTR	P24708 aotus trivi
26	280	4.5	522	1 INVO_HYLLA	P17941 hylobates i
27	280	4.5	1938	1 MYS_AEQIR	P23473 aequipecten
28	279	4.5	1023	1 GLT_DROME	P33438 dirosophila
29	278.5	4.4	678	1 GARP_PLAUF	P13816 plasmodium
30	277	4.4	377	1 NCK1_HUMAN	P16333 homo sapien
31	276.5	4.4	1300	1 DYNA_NEUCR	Q01397 neuropept
32	276	4.4	407	1 M21_STRPY	P50468 streptococc
33	274.5	4.4	2230	1 GOG4_HUMAN	Q13439 homo sapien

34	273.5	4.4	671	1 VINE_HUMAN	Q60504 homo sapien
35	273.5	4.4	857	1 NFW_CHICK	P16053 gallus gall
36	273.5	4.4	1344	1 IF3A_MOUSE	P23116 mus musculu
37	273.5	4.4	1382	1 IF3A_HUMAN	Q14152 homo sapien
38	273	4.4	1976	1 MYHA_BOVIN	Q27991 bos taurus
39	273	4.4	1976	1 MYHA_RAT	Q91100 rattus norv
40	272.5	4.3	501	1 MYSU_RABIT	Q9105 oryctolagus
41	272	4.3	560	1 INVO_PANPA	P14591 pan paniscu
42	272	4.3	1123	1 DC11_DROME	P18169 dirosophila
43	271.5	4.3	2245	1 MYSJ_DICDI	P54697 dicystostel
44	271	4.3	835	1 INVO_PONPY	P14708 pongo pygma
45	271	4.3	1976	1 MYHA_HUMAN	P35580 homo sapien

## ALIGNMENTS

RESULT 1  
ID ITN1\_MOUSE STANDARD; PRT; 1714 AA.  
AC Q920R4; Q9143; 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Intersectin 1 (EH and SH3 domains protein 1).  
GN ITSN1 OR ITSN OR ESE1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OK NCBI\_TaxID=10090;  
RN  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE=99164083; PubMed=10064583;  
RA Sengar A.S., Wang W., Bishop J., Cohen S., Egan S.E.;  
RT "The EH and SH3 domain ESE proteins regulate endocytosis by linking to  
RT dynamin and Eps15.";  
RL EMBO J. 18:1159-1171(1999).  
RN  
RP SEQUENCE OF 966-1714 FROM N.A. (ISOFORM 2), AND SEQUENCE OF 545-599  
RP FROM N.A. (ISOFORMS 1 AND 2).  
RC STRAIN=129/Ola; TISSUE=Spleen;  
RA Skripkina I.Y., Tsyba L.O., Anoprienko O.V., Slavov D., Tassone F.,  
RA Ryndlich A.V., Gardiner K.;  
RT "Mouse homologues of human chromosome 21 genes.";  
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
CC  
CC - FUNCTION: Adapter protein that may provide indirect link between  
CC the endocytic membrane traffic and the actin assembly machinery.  
CC May regulate the formation of clathrin-coated vesicles.  
CC - SUBUNIT: Interacts with dynamin, SNAP-25 and SNAP-23. Clusters  
CC several dynamin in a manner that is regulated by alternative  
CC splicing. Also binds clathrin-associated proteins and other  
CC components of the endocytic machinery, such as N-WASP, Eps15 and  
CC Stecin 2 (By similarity).  
CC  
CC - SUBCELLULAR LOCATION: Cytoplasmic; membrane-associated protein.  
CC Enriched in synaposomes (By similarity).  
CC - ALTERNATIVE PRODUCTS: At least 2 isoforms, 1/Ese1 (shown here)  
CC and 2; are produced by alternative splicing.  
CC - TISSUE SPECIFICITY: Widely expressed. Expressed at high levels in  
CC brain, heart and skeletal muscle.  
CC  
CC - DOMAIN: SH3-3, SH3-4 and SH3-5, but not SH3-1 and SH3-2 domains,  
CC bind to dynamin (By similarity).  
CC  
CC - DOMAIN: The KLERQ domain binds to SNAP-25 and SNAP-23 (By  
CC similarity).  
CC  
CC - MISCELLANEOUS: Overexpression results in the inhibition of the  
CC transferrin uptake and the blockage of the clathrin-mediated  
CC endocytosis.  
CC  
CC - SIMILARITY: CONTAINS 1 C2 DOMAIN.  
CC - SIMILARITY: CONTAINS 1 DBL-HOMOLOGY (DH) DOMAIN.  
CC - SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.  
CC - SIMILARITY: CONTAINS 1 PH DOMAIN.  
CC - SIMILARITY: CONTAINS 5 SH3 DOMAINS.  
CC - SIMILARITY: CONTAINS 2 EH DOMAINS.  
CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AF132481; AAD19749.1; -;  
 DR EMBL; AF132478; AAD19746.1; -;  
 DR EMBL; AF169621; AAD48848.1; -;  
 DR EMBL; AF356517; AAK40228.1; -;  
 DR HSP; P29354; IGFC.  
 DR MGD; MGI:1338069; Itsn.  
 DR InterPro; IPR000008; C2.  
 DR InterPro; IPR002048; EF-hand.  
 DR InterPro; IPR000261; Eps15\_repeat.  
 DR InterPro; IPR001849; PH.  
 DR InterPro; IPR000219; Rhogef.  
 DR InterPro; IPR001452; SH3.  
 DR Pfam; PF00018; SH3; 5.  
 DR Pfam; PF00036; ehand; 3.  
 DR Pfam; PF00168; C2; 1.  
 DR Pfam; PF00169; PH; 1.  
 DR Pfam; PF00621; Rhogef; 1.  
 DR PRINTS; PR00452; SH3DOMAIN.  
 DR PRODOM; PD000066; SH3; 5.  
 DR SMART; SM00239; C2; 1.  
 DR SMART; SM00054; EFh; 2.  
 DR SMART; SM00027; EH; 2.  
 DR SMART; SM00233; PH; 1.  
 DR SMART; SM00325; Rhogef; 1.  
 DR SMART; SM00326; SH3; 5.  
 DR PROSITE; PS50004; C2\_DOMAIN\_2; 1.  
 DR PROSITE; PS50010; DH\_2; 1.  
 DR PROSITE; PS00018; EF\_HAND; 2.  
 DR PROSITE; PS50031; EH; 2.  
 DR PROSITE; PS50003; PH\_DOMAIN; 1.  
 DR PROSITE; PS50002; SH3; 5.  
 DR Endocytosis; SH3 domain; Repeat; Coiled coil; Calcium-binding;  
 KW Alternative splicing.  
 FT DOMAIN 21 109 EH 1.  
 FT CA BIND 66 78 EF-HAND 1 (POTENTIAL).  
 FT DOMAIN 221 310 EH 2.  
 FT CA BIND 267 279 EF-HAND 2 (POTENTIAL).  
 FT DOMAIN 326 702 LYS/LEU/GLU/ARG/GLN-RICH (KLEK).  
 FT DOMAIN 352 662 COILED COIL (POTENTIAL).  
 FT DOMAIN 738 799 SH3 1.  
 FT DOMAIN 906 964 SH3 2.  
 FT DOMAIN 995 1053 SH3 3.  
 FT DOMAIN 1067 1131 SH3 4.  
 FT DOMAIN 1148 1207 SH3 5.  
 FT DOMAIN 1230 1416 DH.  
 FT DOMAIN 1455 1564 PH.  
 FT DOMAIN 1576 1672 C2 DOMAIN.  
 FT DOMAIN 321 324 POLY-SER.  
 FT VARSPIC 1214 1714 MISSING (IN ISOPFORM 2).  
 SQ SEQUENCE 1714 AA; 194284 MW; 4D7AF298397860A7 CRC64;

Query Match 100.0%; Score 6269; DB 1; Length 1714;  
 Best Local Similarity 100.0%; Pred. No. 1e-237;  
 Matches 1213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAQPTTPGSLDVAIVVERAKHDOOPLSLKPIAGITGDDOARNFPPQSGLPQVLAQ 60  
 Db 1 MAQPTTPGSLDVAIVVERAKHDOOPLSLKPIAGITGDDOARNFPPQSGLPQVLAQ 60

QY 61 IVALADNNDRMDQVERSIAMKLIKTLQGYQLPSTLPVWKQOQVVAISSAPAFGIGI 120  
 Db 61 IVALADNNDRMDQVERSIAMKLIKTLQGYQLPSTLPVWKQOQVVAISSAPAFGIGI 120

QY 121 ASMPPLTVAIVPMGSLFVVGMSPLVSSVPPAIVPLANGAPVVIQPLAFAPATWP 180  
 Db 121 ASMPPLTVAIVPMGSLFVVGMSPLVSSVPPAIVPLANGAPVVIQPLAFAPATWP 180

Db 121 ASMPPLTVAIVPMGSLFVVGMSPLVSSVPPAIVPLANGAPVVIQPLAFAPATWP 180

QY 181 KSSFSRSGPQSLWTQKQAKSPVAGAPPAEWAQVSSRLKYROLFNSHDTMSGHL 240  
 Db 181 KSSFSRSGPQSLWTQKQAKSPVAGAPPAEWAQVSSRLKYROLFNSHDTMSGHL 240

QY 241 TGPQARTILMOSSLPQALASIWLSIDIDQCKLTAEFFILAMHLIDVANGQPLPVLP 300  
 Db 241 TGPQARTILMOSSLPQALASIWLSIDIDQCKLTAEFFILAMHLIDVANGQPLPVLP 300

QY 301 PEYIPSPFRVSSGSGMSVSSSVQDRLPEPSSDQPEKKLPVTFPEKKENFERG 360  
 Db 301 PEYIPSPFRVSSGSGMSVSSSVQDRLPEPSSDQPEKKLPVTFPEKKENFERG 360

QY 361 SVELKRRQALLQQRKEERLAQERAEORKEKERERQOEAKQLELEKQERELER 420  
 Db 361 SVELKRRQALLQQRKEERLAQERAEORKEKERERQOEAKQLELEKQERELER 420

QY 421 QREERKKEIREREAKELEERQOLEWERNRQELNQRKKEGTVVLAKRKTEFE 480  
 Db 421 QREERKKEIREREAKELEERQOLEWERNRQELNQRKKEGTVVLAKRKTEFE 480

QY 481 LEALNDKQHLEKQDRIICRLATQREIESTNKSRELRLAEITHLQOOLQESQOMGL 540  
 Db 481 LEALNDKQHLEKQDRIICRLATQREIESTNKSRELRLAEITHLQOOLQESQOMGL 540

QY 541 IPEKOILSDQKQVQNSLHRDSLITLKRALAEKELARQOLREOLDEVERETRSKLOEID 600  
 Db 541 IPEKOILSDQKQVQNSLHRDSLITLKRALAEKELARQOLREOLDEVERETRSKLOEID 600

QY 601 VFNNOQLKELREIHSKQOLQOKRSLEAARLKQKEERKSLLEKQEKEDQORVQSRDQWL 660  
 Db 601 VFNNOQLKELREIHSKQOLQOKRSLEAARLKQKEERKSLLEKQEKEDQORVQSRDQWL 660

QY 661 EHVOQEOGRPRKPHEDRLRKEDSVRKEAEERAKPEMODQKSLFPHQEPKALATQA 720  
 Db 661 EHVOQEOGRPRKPHEDRLRKEDSVRKEAEERAKPEMODQKSLFPHQEPKALATQA 720

QY 721 PWSTTEKGLTISAQESKVVYRRLYPFESRSHDEITIQGDIIVMDESQTGPGMG 780  
 Db 721 PWSTTEKGLTISAQESKVVYRRLYPFESRSHDEITIQGDIIVMDESQTGPGMG 780

QY 781 ELKGTGMPANVAKIENEPVTPAKPVTDLTSAAPKALARETPALPTSSSEPTTP 840  
 Db 781 ELKGTGMPANVAKIENEPVTPAKPVTDLTSAAPKALARETPALPTSSSEPTTP 840

QY 841 NNWADPSSSTWPSSENEKEPTDNDTMAAQPSLTVPSAGQLFQSAFTPATATGSSPSV 900  
 Db 841 NNWADPSSSTWPSSENEKEPTDNDTMAAQPSLTVPSAGQLFQSAFTPATATGSSPSV 900

QY 901 GQGEVVEGLQAOALYPMRAKKDNHLENFKSDVITVLEQDDMMWGEVQGGKMPKSVK 960  
 Db 901 GQGEVVEGLQAOALYPMRAKKDNHLENFKSDVITVLEQDDMMWGEVQGGKMPKSVK 960

QY 961 LISGPRKSTSIDTPTSPASLKRVASPAKPAIPGEFTIAMYTYSSEOGDLTFQOGD 1020  
 Db 961 LISGPRKSTSIDTPTSPASLKRVASPAKPAIPGEFTIAMYTYSSEOGDLTFQOGD 1020

QY 1021 VIVVTKKQDMMWTVGVGKSGVFPBNYRLKDSGSGTAGTGSIGKKPEIAQVIASTAA 1080  
 Db 1021 VIVVTKKQDMMWTVGVGKSGVFPBNYRLKDSGSGTAGTGSIGKKPEIAQVIASTAA 1080

QY 1081 TGPQOLTLAPGQLIRKKNPGMWEGELQARGKKRQIGWFPANVVKLLSGTSKITPTE 1140  
 Db 1081 TGPQOLTLAPGQLIRKKNPGMWEGELQARGKKRQIGWFPANVVKLLSGTSKITPTE 1140

QY 1141 LPRTAQVPAVCVIGMYDTAQNDELAFSGQIINVLNKDPPMWKGEVSGQVGLFPSPN 1200  
 Db 1141 LPRTAQVPAVCVIGMYDTAQNDELAFSGQIINVLNKDPPMWKGEVSGQVGLFPSPN 1200

QY 1201 VYKLTITDMDPSQO 1213  
 Db 1201 VYKLTITDMDPSQO 1213

RESULT 2  
ID ITN1 RAT STANDARD; PRT; 1217 AA.  
AC O9WVE9; O9WVE1.  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Intersectin 1 (EH domain and SH3 domain regulator of endocytosis 1).  
GN ITSN1 OR ITSN OR ESHS1.  
OS Rattus norvegicus (Rat).  
OC Bakaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OC NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RC TISSUE=Brain;  
RX MEDLINE=9303609; PubMed=10373452;  
RA Okamoto M., Schoch S., Suedhof T.C.;  
RT "HSH1/Intersectin, a protein that contains EH and SH3 domains and  
RT binds to dynamin and SNAP-25. A protein connection between exocytosis  
RT and endocytosis?";  
RT J. Biol. Chem. 274:18446-18454(1999).  
RU  
CC - FUNCTION: Adapter protein that may provide indirect link between  
CC the endocytic membrane traffic and the actin assembly machinery.  
CC May regulate the formation of clathrin-coated vesicles.  
CC - SUBUNIT: Interacts with dynamin, SNAP-25 and SNAP-23. Clusters  
CC several dynamin in a manner that is regulated by alternative  
CC splicing. Also binds clathrin-associated proteins and other  
CC components of the endocytic machinery, such as N-WASP, Eps15 and  
CC Stomatin 2 (by similarity).  
CC - SUBCELLULAR LOCATION: Cytoplasmic; membrane-associated protein.  
CC Enriched in synaposomes.  
CC - ALTERNATIVE PRODUCTS: At least 2 isoforms: 1 (shown here) and 2;  
CC are produced by alternative splicing.  
CC - TISSUE SPECIFICITY: Highly expressed in brain.  
CC - DOMAIN: SH3-3, SH3-4 and SH3-5, but not SH3-1 and SH3-2 domains,  
CC bind to dynamin.  
CC - DOMAIN: The KLERQ domain binds to SNAP-25 and SNAP-23.  
CC - SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.  
CC - SIMILARITY: CONTAINS 5 SH3 DOMAINS.  
CC - SIMILARITY: CONTAINS 2 EH DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: ARI27798; AAD30271.1; -;  
DR EMBL: ARI32672; AAD31026.1; -;  
DR HSSP: P29354; IGFC.  
DR InterPro: IPR002048; EF-hand.  
DR InterPro: IPR000261; EPHS1\_repeat.  
DR InterPro: IPR001452; SH3.  
DR Pfam: PF00018; SH3; 5.  
DR Pfam: PF00016; ehand; 3.  
DR PRINTS: PR00452; SH3DOMAIN.  
DR PRODOM: PD000066; SH3; 5.  
DR SMART: SM00054; EPH; 2.  
DR SMART: SM00027; EH; 2.  
DR SMART: SM00326; SH3; 5.  
DR PROSITE: PS00018; EF\_HAND; 2.  
DR PROSITE: PS50031; EH; 2.  
DR PROSITE: PS50002; SH3; 5.  
KM Endocytosis; SH3 domain; Repeat; Coiled coil; Calcium-binding;  
KM Alternative splicing.  
FT CA BIND 21 109 EH 1.  
FT 66 78 EF-HAND 1 (POTENTIAL).  
FT DOMAIN 21 109  
FT DOMAIN 221 310 EH 2.

FT CA BIND 267 279 EF-HAND 2 (POTENTIAL).  
FT DOMAIN 326 702 LYS/LEU/ARG/GLN-RICH (KLERQ).  
FT DOMAIN 350 670 COILED COIL (POTENTIAL).  
FT DOMAIN 737 803 SH3 1.  
FT DOMAIN 910 968 SH3 2.  
FT DOMAIN 999 1057 SH3 3.  
FT DOMAIN 1071 1135 SH3 4.  
FT DOMAIN 1152 1211 SH3 5.  
FT DOMAIN 321 324 SH3 5.  
FT VARSPLIC 1003 1073 POLY-SER.  
FT SEQUENCE 1217 AA; 137154 MW; 6C13238A5E5B34B CRC64;  
Query Match 96.0%; Score 6021; DB 1; Length 1217;  
Best Local Similarity 95.7%; Pred. No. 3,3e-228;  
Matches 1166; Conservative 19; Mismatches 27; Indels 6; Gaps 2;  
QY 1 MAQFPFPFGSLIDVAIVTEERAKHDQFLSKIPAGITGDOANRFFFGSLPQVLAQ 60  
DB 1 MAQFPFPFGSLIDVAIVTEERAKHDQFQSLKPISGFITDQANRFFFGSLPQVLAQ 60  
QY 61 IVALDMNNDGMDQVERSIAMKLIKLGQYQLPSTLPVWKQOPVAISSAPAFGIGI 120  
DB 61 IVALDMNNDGMDQVERSIAMKLIKLGQYQLPSTLPVWKQOPVAISSAPAFGIGI 120  
QY 121 ASMPPLTAVAPVPMGSIPIVGMSPPLVSSVPAAVPLANGAPVYIQPLPAFAPATWP 180  
DB 121 ASMPPLTAVAPVPMGSIPIVGMSPPLVSSVPAAVPLANGAPVYIQPLPAFAPATWP 180  
QY 181 KSSSFRSGSGSLNTKIQKASFPVVASAPPAEAVVQSSRLKTKROLFNSHDKTMSHL 240  
DB 181 KSSSFRSGSGSLNTKIQKASFPVVASAPPAEAVVQSSRLKTKROLFNSHDKTMSHL 240  
QY 241 TGPQARTIMQSSLPQAOASIMNLSIDIDODGKLTAEFFILAMHLIDVAMSGQPLPVLP 300  
DB 241 TGPQARTIMQSSLPQAOASIMNLSIDIDODGKLTAEFFILAMHLIDVAMSGQPLPVLP 300  
QY 301 PEYIPPSFRVRSVSGMSVSISSSVQDRLPEPSSSEDDQPEKKLPVTFEDKKENFERG 360  
DB 301 PEYIPPSFRVRSVSGMSVSISSSVQDRLPEPSSSEDDQPEKKLPVTFEDKKENFERG 360  
QY 361 SVELKKRRALLEQQRKQERLAQLEAEQERKEPERBOEAKQOLEKOLEKORELER 420  
DB 361 SVELKKRRALLEQQRKQERLAQLEAEQERKEPERBOEAKQOLEKOLEKORELER 420  
QY 361 NLEKRRRALLEQQRKQERLAQLEAEQERKEPERBOEAKQOLEKOLEKORELER 420  
DB 361 NLEKRRRALLEQQRKQERLAQLEAEQERKEPERBOEAKQOLEKOLEKORELER 420  
QY 421 QSEERKREIEREAKKELEEROLEEMERNROBLNQRKKEOGTVLAKRKTLFE 480  
DB 421 QSEERKREIEREAKKELEEROLEEMERNROBLNQRKKEOGTVLAKRKTLFE 480  
QY 481 LEALNDKQHOLEGKLQDRCRLATQROEIESTNKRRELRIAEITHLQOQLOESQMLGRL 540  
DB 481 LEALNDKQHOLEGKLQDRCRLATQROEIESTNKRRELRIAEITHLQOQLOESQMLGRL 540  
QY 541 IPEKOTLSDQLKQVQNSLHRDSLTLTKRALEAKELAQOLREOLDEVERETRSKLOEID 600  
DB 541 IPEKOTLSDQLKQVQNSLHRDSLTLTKRALEAKELAQOLREOLDEVERETRSKLOEID 600  
QY 601 VNNOLKELREIHSKQOLQKORSLEAARLKQKEERKSLLEKQEKEDQORVORBDKWL 660  
DB 601 VNNOLKELREIHSKQOLQKORSLEAARLKQKEERKSLLEKQEKEDQORVORBDKWL 660  
QY 661 EHVQOEQPRPKPHEBDRLKREDSVRKKEAEERAKPEMDQKSGRLFPHPQEPKLAQ 720  
DB 661 EHVQOEQPRPKPHEBDRLKREDSVRKKEAEERAKPEMDQKSGRLFPHPQEPKLAQ 720  
QY 721 PWTTEKGPPLTISAQESAKVYVYRALYFPESRSHDEITIQGDIVMKGWVDSQTGE 779  
DB 721 PWTTEKGPPLTISAQESAKVYVYRALYFPESRSHDEITIQGDIVMKGWVDSQTGE 779  
QY 776 GVLGELGKGTGWPANAAEKIPENEVPTPAKPYTDLUSAPAPKLAJRETPAPVPSSE 835  
DB 776 GVLGELGKGTGWPANAAEKIPENEVPTPAKPYTDLUSAPAPKLAJRETPAPVPSSE 835  
QY 780 GVLGEPKGTGWPANAAEKIPENEVPTPAKPYTDLUSAPAPKLAJRETPAPVPSSE 839  
DB 780 GVLGEPKGTGWPANAAEKIPENEVPTPAKPYTDLUSAPAPKLAJRETPAPVPSSE 839  
QY 836 PSTTNNMADPSSSTWSSSNEKPEPTDNDWDTAAQPSLTVPASAGLRRORSAPTPATATGSS 895  
DB 836 PSTTNNMADPSSSTWSSSNEKPEPTDNDWDTAAQPSLTVPASAGLRRORSAPTPATATGSS 895



```

DR InterPro: IPR000261; ERF5_repeat.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000219; RhogEF.
DR InterPro: IPR001452; SH3.
DR Pfam: PF000018; SH3; 5.
DR Pfam: PF000036; ehfand; 3.
DR Pfam: PF00168; C2; 1.
DR Pfam: PF00169; PH; 1.
DR Pfam: PF00621; RhogEF; 1.
DR PRINTS: PR00452; SH3DOMAIN.
DR ProDom: PD000066; SH3; 4.
DR SMART: SM00326; SH3; 4.
DR PROSITE: PS00004; C2_DOMAIN_2; 1.
DR PROSITE: PS00010; DH_2; 1.
DR PROSITE: PS00018; EF_HAND; 2.
DR PROSITE: PS00031; EH; 2.
DR PROSITE: PS00033; PH_DOMAIN; 1.
DR PROSITE: PS0002; SH3; 5.
DR Endocytosis; SH3 domain; Repeat; Coiled coil; Calcium-binding;
KW Alternative splicing.
FT DOMAIN 21 109 EH 1.
FT CA_BIND 66 77 EF_HAND 1 (POTENTIAL).
FT DOMAIN 221 310 EH 2.
FT CA_BIND 267 278 EF_HAND 2 (POTENTIAL).
FT DOMAIN 326 702 LYS/LEU/GLU/ARG/GLN-RICH (KIERO).
FT DOMAIN 351 705 COILED COIL (POTENTIAL).
FT DOMAIN 740 806 SH3 1.
FT DOMAIN 913 971 SH3 2.
FT DOMAIN 1002 1060 SH3 3.
FT DOMAIN 1074 1138 SH3 4.
FT DOMAIN 1155 1214 SH3 5.
FT DOMAIN 1237 1423 DH.
FT DOMAIN 1462 1571 PH.
FT DOMAIN 1583 1679 C2_DOMAIN.
FT VARSPPLIC 770 MISSING (IN ISOFORM 3).
FT VARSPPLIC 1006 1076 MISSING (IN ISOFORM 3 AND ISOFORM 4).
FT CONFLICT 114 114 P -> A (IN REF. 2).
FT CONFLICT 1088 1088 T -> A (IN REF. 3).
FT CONFLICT 1109 1109 G -> R (IN REF. 3).
FT CONFLICT 1137 1137 S -> N (IN REF. 1).
FT CONFLICT 1361 1361 E -> A (IN REF. 2).
FT CONFLICT 1474 1474 N -> S (IN REF. 2).
SQ SEQUENCE 1721 AA; 195532 MW; 8189DDA0F145B4B5 CRC64;

Query Match 93.2%; Score 5842.5; DB 1; Length 1721;
Best Local Similarity 92.5%; Pred. No. 4.3e-221;
Matches 1128; Conservative 39; Mismatches 46; Indels 7; Gaps 3;

```

```

Qy 360 GSVLEKRRQALLLEQQRKEQRLAQLEAEQERKEREQOEQAEQRLQLEKQREBLE 419
Db 361 GNLELEKRRQALLLEQQRKEQRLAQLEAEQERKEREQOEQAEQRLQLEKQREBLE 420
Qy 420 RQREERREKELERRAARELERQRLQLEAEQERKEREQOEQAEQRLQLEKQREBLE 479
Db 421 RQREERREKELERRAARELERQRLQLEAEQERKEREQOEQAEQRLQLEKQREBLE 480
Qy 480 ELEALNDKKHQLQGLQDRIQRLATQROEISTYKSRRLAEITTHLQOQOEQOMLGR 539
Db 481 ELEALNDKKHQLQGLQDRIQRLATQROEISTYKSRRLAEITTHLQOQOEQOMLGR 540
Qy 540 LIPEKQILSDQKQVQNSLHRDSLTLTKRALAEKELARQQLREQLDEVERETSKLQEI 599
Db 541 LIPEKQILSDQKQVQNSLHRDSLTLTKRALAEKELARQQLREQLDEVERETSKLQEI 600
Qy 600 DVFNQQLKELREIHSKQQLQQRSLAEARLQKQERKSLLEKQEKEDAQRVVERDKQW 659
Db 601 DVFNQQLKELREIHSKQQLQQRSLAEARLQKQERKSLLEKQEKEDAQRVVERDKQW 660
Qy 660 LEHVQOE-EQPRPKRHEBRLKREDSYRKKEAEERAKPEQODKOSRLFPHQBPALAT 718
Db 661 LEHVQOEDEHQPRKLEHEERLKEBESYKKGDSGEKQOEADKGLFPHQBPAPAV 720
Qy 719 QAPWSTTEKGLTISAQSVKVVYRALYPRESSHDEITIQPDIVY----VDESQTG 773
Db 721 QAPWSTTEKGLTISAQSVKVVYRALYPRESSHDEITIQPDIVYVWKGWVDESQTG 780
Qy 774 EPGWLIGELKGTGFWFPANVAEKIPENEVPTPAKPVYDLSAPAPKALARTPAPLVTS 833
Db 781 EPGWLIGELKGTGFWFPANVAEKIPENEVPAKPVYDLSAPAPKALARTPAPLVTS 840
Qy 834 SEPTTPNNMADFSTWSSSNEKPEETDMDTMAOPLTVPSAGOLRORSAFPTATG 893
Db 841 SEPTTPNNMADFSTWSSSNEKPEETDMDTMAOPLTVPSAGOLRORSAFPTATG 900
Qy 894 SSPSPVLQGEKEVEGLQALYPRAKKNDLNFNKSDVITVLEQODMMWGEVQGGKW 953
Db 901 SSPSPVLQGEKEVEGLQALYPRAKKNDLNFNKSDVITVLEQODMMWGEVQGGKW 960
Qy 954 PPKSVYKLIISGPRKSTSIDTGPTEPSPSLKRVASPAKPAIPBEETIANTYSSSGCD 1013
Db 961 PPKSVYKLIISGPRKSTSIDTGPTEPSPSLKRVASPAKPAIPBEETIANTYSSSGCD 1020
Qy 1014 LTFQOGDVIYVTKKDGDMWGTGVPKSGVPSNVYRLKDSGSGTAGTSGLGKKPBLAQ 1073
Db 1021 LTFQOGDVIYVTKKDGDMWGTGVPKSGVPSNVYRLKDSGSGTAGTSGLGKKPBLAQ 1080
Qy 1074 VIASVYATGPEQLTLAPQQLILIRKKNPGWMEGELQARGKKRQIGFPPANYVVLSPGT 1133
Db 1081 VIASVYATGPEQLTLAPQQLILIRKKNPGWMEGELQARGKKRQIGFPPANYVVLSPGT 1140
Qy 1134 SKITTELPKTAQVAVQVIGMTDYTAQNDDELAFSGKQIINVLAKEDPMMKGEVSGQ 1193
Db 1141 SKITTELPKSTALAAQVIGMTDYTAQNDDELAFSGKQIINVLAKEDPMMKGEVSGQ 1200
Qy 1194 VGLFPSNVYKLTMDPSQ 1213
Db 1201 VGLFPSNVYKLTMDPSQ 1220

RESULT 4
ITN1_XENLA STANDARD; PRT: 1270 AA.
AC 0422877
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Intersectin 1.
CN ITSN1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;

```



OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP TRISUPE-Oocyte;  
 RA MEDLINE=99030416; PubMed=9813051;  
 RX Yamabhai M., Hoffman N.G., Hardison N.L., McPherson P.S.,  
 RA Castagnoli L., Cesarini G., Kay B.K.;  
 RT "intersecin, a novel adaptor protein with two eps15 homology and five  
 RL src homology 3 domains";  
 J. Biol. Chem. 273:31401-31407(1998).  
 CC -!- FUNCTION: Adaptor protein that may provide indirect link between  
 CC the endocytic membrane traffic and the actin assembly machinery.  
 CC May regulate the formation of clathrin-coated vesicles.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic; membrane-associated protein.  
 CC -!- ERITCHED IN SYNAPSOSES (By similarity).  
 CC -!- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 5 SH3 DOMAINS.  
 CC -!- SIMILARITY: CONTAINS 2 EH DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF032118; AAC73068.1; -;  
 DR HSSP: P29355; 1SEM.  
 DR InterPro: IPR002048; EF-hand.  
 DR InterPro: IPR000261; EPS15\_repeat.  
 DR InterPro: IPR001452; SH3.  
 DR Pfam: PF00018; SH3; 5.  
 DR Pfam: PF00036; ehfand; 2.  
 DR PRINTS: PR00452; SH3DOMAIN.  
 DR PRODOM: PD000066; SH3; 5.  
 DR SMART: SM00054; Efh; 2.  
 DR SMART: SM00027; EH; 2.  
 DR SMART: SM00326; SH3; 5.  
 DR PROSITE: PS00018; EF\_HAND; 2.  
 DR PROSITE: PS50031; EH; 2.  
 DR PROSITE: PS50002; SH3; 5.  
 KM Endocytosis; SH3 domain; Repeat; Coiled coil; Calcium-binding.  
 FT DOMAIN 21 109  
 FT CA\_BIND 66 77 EF\_HAND 1 (POTENTIAL).  
 FT DOMAIN 220 309  
 FT CA\_BIND 266 277 EF\_HAND 2 (POTENTIAL).  
 FT DOMAIN 325 697 LYS/LEU/GLU/ARG/GLN-RICH (KLEERQ).  
 FT DOMAIN 349 691 COILED COIL (POTENTIAL).  
 FT DOMAIN 732 793 SH3 1.  
 FT DOMAIN 897 955 SH3 2.  
 FT DOMAIN 986 1044 SH3 3.  
 FT DOMAIN 1058 1122 SH3 4.  
 FT DOMAIN 1139 1198 SH3 5.  
 SQ SEQUENCE 1270 AA; 143670 MW; EA940C1F6B6A858 CRC64;  
 Query Match 80.8%; Score 5063.5; DB 1; Length 1270;  
 Best Local Similarity 80.6%; Pred. No. 7.6e-191;  
 Matches 980; Conservative 98; Mismatches 123; Indels 15; Gaps 8;

QY 179 MPKSSFSRSGSGSLNTKLOKASFDVASAPPAEAWPQSSRLTKYROLFNSHDKTMSG 238  
 DB 178 LKSSSFGSRVAGQINTKLOKASFDVASAPPAEAWPQSSRLTKYROLFNSHDKTMSG 237  
 QY 239 HLTGPQARTIMQSSLPQALASIMNLSIDIDQDGLTAEEFIILAMHLIDVAMSGQPLPV 298  
 DB 238 NLTGPQARTIMQSSLPQALASIMNLSIDIDQDGLTAEEFIILAMHLIDVAMSGQPLPV 297  
 QY 299 LPPEYIPSPFRVSGSGMSYSSSDVDQRLPEPSSSEDEQPEPKLPVTFEDKKRNF 358  
 DB 298 LPPEYIPSPFRVSGSGMSYSSSDVDQRLPEPSSSEDEQPEPKLPVTFEDKKRNF 357  
 QY 359 RGSVLEKRRQALAEQKEQERLAQLEAEQEKERERQEAQROLELEKQEREL 418  
 DB 358 RNLBLEKRRQALAEQKEQERLAQLEAEQEKERERQEAQROLELEKQEREL 417  
 QY 419 EKOREEERKEIEREAKRELERQOLEWERNRQELNQRNKEQEGTVLKKARRTLE 478  
 DB 418 EKOREEERKEIEREAKRELERQOLEWERNRQELNQRNKEQEGTVLKKARRTLE 477  
 QY 479 FELELNKQQLBEKLODIFCRLATQROELIESTNKSRELIAETHLQOOLQSSQMLG 538  
 DB 478 FELELNKQQLBEKLODIFCRLATQROELIESTNKSRELIAETHLQOOLQSSQMLG 537  
 QY 539 RLPEKQILSDQLKQVQNSLHRSDILTLKRALBAKELARQOLREQDEVERETRSKLOE 598  
 DB 538 RLPEKQILSDQLKQVQNSLHRSDILTLKRALBAKELARQOLREQDEVERETRSKLOE 597  
 QY 599 IDVFNNQLEKELEIHSKOOLKORSLEAARLKQKOEKSLLEKQEDNORRQOEKDK 658  
 DB 598 IDVFNNQLEKELEIHSKOOLKORSLEAARLKQKOEKSLLEKQEDNORRQOEKDK 657  
 QY 659 WLEHYQOEOPRPRKPHEDRLKREDSYRKKAEKABERAKEMQDKOSRLFPHQBPALAT 718  
 DB 658 WQDRVQKEBE--RYKFOEBEKKREESYQKCEVEK--KPELOEKPNKPFHQPPRGKLG 713  
 QY 719 QAPWSTTEKGPLTISAQESVKKVYVYRALPYESSSHDEITTPQDVIWVDESQTEBGM 778  
 DB 714 QIPMNTEKAPLTIN-QGDVKKVYVYRALPYEDARSHDEITTEPDIMVDESQTEBGM 772  
 QY 779 GGELEKGTGMPANYAEKIPENNEVPTPAKPYTDLTSAAPKLAREPAPL-PYTSSEPS 837  
 DB 773 GGELEKGTGMPANYAEKIPENNEVPTPAKPYTDLTSAAPKLAREPAPL-PYTSSEPS 828  
 QY 838 TTPNNMADPSSSTWPSSENEKETDMDTMAAOPSLTVPSAQOLRORSAFTPATATGSSPS 897  
 DB 829 TNSNNMADPSSSTWPSSENEKETDMDTMAAOPSLTVPSAQOLRORSAFTPATATGSSPS 888  
 QY 898 PVLGGQEKYEGLOAQALYPMRAKKDNHLNFKNSDYITVLEQDDMMWREGVQCGKMPFKS 957  
 DB 889 PVLGGQEKYEGLOAQALYPMRAKKDNHLNFKNSDYITVLEQDDMMWREGVQCGKMPFKS 948  
 QY 958 YKYLISGPRKSTSIDTPTSPASLKRVASPAKPAIPGEFTIAMTYESESQGDLLTFQ 1017  
 DB 949 YKYLISGPRKSTSIDTPTSPASLKRVASPAKPAIPGEFTIAMTYESESQGDLLTFQ 1008  
 QY 1018 QGDVIVTKKQDQMTGTGDKSGVFPNSNYVRLKDSSEGSTAGTGTGSGKPELIAQVIA 1077  
 DB 1009 QGDVIVTKKQDQMTGTGDKSGVFPNSNYVRLKDSSEGSTAGTGTGSGKPELIAQVIA 1068  
 QY 1078 YAATPEQTLTAPQGLILIRKKNPGWMEGELQAKGKRQIGWPPANVVKLISGTSKIT 1137  
 DB 1069 YAATPEQTLTAPQGLILIRKKNPGWMEGELQAKGKRQIGWPPANVVKLISGTSKIT 1128  
 QY 1138 PTELEPRTAVQAPVQVIGVYDTAQNDELAFSKQOINVLNKEDPDWKKKEVSGQVGLF 1197  
 DB 1129 PTELEPRTAVQAPVQVIGVYDTAQNDELAFSKQOINVLNKEDPDWKKKEVSGQVGLF 1188  
 QY 1198 PSNYVKLTITMDPSSQ 1213  
 DB 1189 PSNYVKLTITMDPSSQ 1204

RESULT 5  
ID ITN2\_HUMAN STANDARD; PRT; 1696 AA.  
AC O9NZM3; O9NZG0; O9ULG4; O9NZM2; O9HAK4; O9NXB6; O9S062; Q15812;  
AD 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DE Intersectin 2 (SH3 domain-containing protein 1B) (SH3P18) (SH3P18-like  
WASP associated protein).  
GN ITSN2 OR SH3D18 OR SWAP OR KIAA1256.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 3).  
RC TISSUE=Fetal brain, and fetal liver;  
RX MEDLINE=20382722; PubMed=10922467;  
RA Pucharcos C., Estivill X., de la Luna S.;  
RT "Intersectin 2, a new multimodular protein involved in clathrin-  
mediated endocytosis.";  
RL FEBS Lett. 478:43-51(2000).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 2).  
RA Seifert M., Engel M., Welter C.;  
RT "Intersectin 2 (SH3D18), human homolog of mouse Ees2 protein.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 2).  
RC TISSUE=Brain;  
RX MEDLINE=20039619; PubMed=10574462;  
RA Nagase T., Ishikawa K.-I., Kikuno R., Hirokawa M., Nomura N.,  
RA Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XV.  
The complete sequences of 100 new cDNA clones from brain which code  
for large proteins in vitro.";  
RL DNA Res. 6:337-345(1999).  
RN [4]  
RP SEQUENCE OF 258-720 FROM N.A. (ISOFORM 2), AND SEQUENCE OF 1261-1696  
FROM N.A.  
RC TISSUE=Embryo;  
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,  
RA Tanase T., Nomura Y., Togiya S., Komai R., Takeuchi K.,  
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,  
RA Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y., Oshima A.;  
RT "NEO human cDNA sequencing project.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE OF 821-1284 FROM N.A. (ISOFORM 1).  
RA Wilson L.A., Fields D., Cruz L., Friesen J., Siminovich K.A.;  
RT "SH3P18-like wasp associated protein (SWAP): A multiple SH3 domain  
containing protein that interacts with WASP.";  
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE OF 944-1191 FROM N.A. (ISOFORM 4).  
RC TISSUE=Bone marrow;  
RX MEDLINE=98294438; PubMed=9630982;  
RA Sparks A.B., Hoffman N.G., McConnell S.J., Fowlkes D.M., Kay B.K.;  
RT "Cloning of ligand targets: systematic isolation of SH3 domain-  
containing proteins.";  
RL Nat. Biotechnol. 14:741-744(1996).  
RN [7]  
RP ALTERNATIVE SPLICING.  
RC TISSUE=Fetal liver, and Brain;  
RX MEDLINE=21548828; PubMed=11690630;  
RA Pucharcos C., Casas C., Nadal M., Estivill X., de la Luna S.;  
RT "The human intersectin genes and their spliced variants are  
differentially expressed.";  
RL Biochim. Biophys. Acta 1521:1-11(2001).  
CC -!- FUNCTION: Adapter protein that may provide indirect link between  
the endocytic membrane traffic and the actin assembly machinery.  
CC May regulate the formation of clathrin-coated vesicles.

CC -!- SUBUNIT: Belongs to a complex that may contain multimers of ITSN1,  
CC ITSN2 and Eps15, and different partners according to the step in  
CC the endocytic process.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- ALTERNATIVE PRODUCTS: At least 4 isoforms; 1/ITSN2-L (shown here),  
CC 2, 3/ITSN2-S1 and 4/ITSN2-S2/SH3P18; are produced by alternative  
CC splicing.  
CC -!- TISSUE SPECIFICITY: Ubiquitous. Isoform 1 is primarily expressed  
CC in adult heart and liver.  
CC -!- MISCELLANEOUS: Overexpression results in the inhibition of the  
CC transferrin uptake and the blockage of the clathrin-mediated  
CC endocytosis.  
CC -!- SIMILARITY: CONTAINS 1 C2 DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY (DH) DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 EF-HAND CALCIUM-BINDING DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.  
CC -!- SIMILARITY: CONTAINS 5 SH3 DOMAINS.  
CC -!- SIMILARITY: CONTAINS 2 EH DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF182198; AAF59903.1; -;  
DR EMBL; AF182199; AAF59904.1; -;  
DR EMBL; AF248540; AAF63600.1; ALT\_INIT.  
DR EMBL; AB033082; BAA86570.1; ALT\_INIT.  
DR EMBL; AB021545; BAB13841.1; -;  
DR EMBL; AK000302; BAA91068.1; -;  
DR EMBL; AF001630; AAD00899.1; -;  
DR EMBL; U61167; AAC50593.1; -;  
DR HSSP; P29355; 1SEM.  
DR Genew; HGNC:6184; ITSN2.  
DR MIM; 604464; -;  
DR InterPro; IPR000008; C2.  
DR InterPro; IPR002048; EFS15\_repeat.  
DR InterPro; IPR000261; EFS15\_repeat.  
DR InterPro; IPR001849; PH.  
DR InterPro; IPR000219; RhGEF.  
DR InterPro; IPR001452; SH3.  
DR Pfam; PF00018; SH3; 5.  
DR Pfam; PF00168; C2; 1.  
DR Pfam; PF00169; PH; 1.  
DR Pfam; PF00621; RhGEF; 1.  
DR PRINTS; PR00499; P67PHOX.  
DR PRINTS; PR00452; SH3DOMAIN.  
DR SMART; SM00239; C2; 1.  
DR SMART; SM0054; Efh; 2.  
DR SMART; SM00027; EH; 2.  
DR SMART; SM00233; PH; 1.  
DR SMART; SM00325; RhGEF; 1.  
DR SMART; SM00326; SH3; 5.  
DR PROSITE; PS50004; C2 DOMAIN\_2; 1.  
DR PROSITE; PS50010; DH\_2; 1.  
DR PROSITE; PS00018; EF\_HAND; 1.  
DR PROSITE; PS50003; PH DOMAIN; 1.  
DR PROSITE; PS50002; SH3; 5.  
DR Endocytosis; SH3 domain; Repeat; Coiled coil; Calcium-binding;  
KW Alternative splicing.  
FT CA\_BIND 21..109  
FT CA\_BIND 66..78  
FT DOMAIN 243..332  
FT DOMAIN 360..755  
FT DOMAIN 756..817  
FT DOMAIN 897..955  
FT DOMAIN 980..1038  
FT DOMAIN 1052..1116  
FT DOMAIN 1126..1185  
FT EF\_HAND (POTENTIAL).  
FT COILED COIL (POTENTIAL).  
FT SH3 1.  
FT SH3 2.  
FT SH3 3.  
FT SH3 4.  
FT SH3 5.

```

FT  DOMAIN  1208  1394  DH.
FT  DOMAIN  1433  1543  PH.
FT  DOMAIN  1555  1651  C2 DOMAIN.
FT  VARSLIC  621  647  MISSING (IN ISOFORM 2).
FT  VARSLIC  1234  1248  FORMASGFLTEGE -> WRLLASSRGICCLS (IN
FT  VARSLIC  1249  1696  ISOFORM 3).
FT  VARSLIC  1192  1696  MISSING (IN ISOFORM 3).
FT  CONFLICT  290  290  MISSING (IN ISOFORM 4).
FT  CONFLICT  678  678  V -> I (IN REF. 4).
FT  CONFLICT  822  827  R -> G (IN REF. 4).
FT  CONFLICT  944  950  KAVSPK -> FFAAST (IN REF. 5).
FT  CONFLICT  1278  1284  WEPKSYV -> EFAAST (IN REF. 6).
FT  CONFLICT  1552  1552  GEMKPYQ -> VDAANS (IN REF. 5).
FT  CONFLICT  1552  1552  K -> Q (IN REF. 2 AND 3).
SQ  SEQUENCE  1696 AA; 193329 MW; 17B5C8629B8CF9B CRC64;

Query Match 48.9%; Score 3068; DB 1; Length 1696;
Best Local Similarity 51.0%; Pred. No. 6,8e-113;
Matches 656; Conservative 187; Mismatches 274; Indels 170; Gaps 30;

1 MAOPPTFGSLDVAITVEERAKHDQFLSKPIAGITGDARNFFQSGLPQPVLAQ 60
1 MAOPPTAMNGPMNMTTSEERTKHDQFDMKPSGYITIGDARNFFLOSGLPAPVLA 60
61 IMAIADNNDGRMDQVEFSIAMKLIKIKLOGYQLPSTLPVPMKQOQV -AISAPAFIG 118
61 IMAISDINKGKMDQEFISIAMKLIKIKLOGYQLPVLPIPMKQPMFSPPLISA-RFGMG 119
119 GIAMSPPLTAVAPVPMG-----SIPVGMSPPLVSSVPPAAVPLANGAPV 165
120 ---SMPVLSTIPQLPPAPAITSLSSATSGTNLPLMPTLVPVSIVSSISPL---NGIASL 173
166 IOPLPAFAHPATWPKSSSES--RSGPGSOLNTKLQKAGS-FDVASAP----- 211
174 IQPLP-IPYSSSTLPHGSSYSILMWGFG--GASICKAQSLIDIGSSSSSTSTASLSGNS 229
212 ----AAEWAHPQSSRLKRYROLFNSHDKTMGHLTGPOARTILMOSSLPQAOLASINMLSP 267
230 PKTGTSMAVQPTRLKRYROKFNLTDSKMSGYISGFQARNALLOSLSQTQLATIMTLAD 289
268 IDODGKLTAEFFILAMHLIDVAMSGOPLPVLPEVYIPPSFRVRSAGSVISSSSVDQ 327
290 VDSGOLKAEFFILAMHLIDVAMSGOPLPVLPEVYIPPSFRVRSAGSVISSSSVDQ 341
328 RLPESSEDEQPEKLPVTFEDKKBENFERGVEIEKRRQALLLEQKREOERLAQER 387
342 TLPSYQKMQ--EEBPQKLPVTFEDKKBENFERGVEIEKRRQALLLEQKREOERLAQER 400
388 AEOERKERERQEOAKQOLELEKOLEKOREROREERREKREIRRAAKRELERQOLE 447
401 EEWKRORELOEOWKOLELEKOLEKOREROREERREKREIRRAAKRELERQOLE 460
448 WERNROELLINORKKEOGTVLAKRKTELELEALINDKKHOLEGLKODIRCLATQRO 507
461 WERIRROELLINORKKEOGTVLAKRKTELELEALINDKKHOLEGLKODIRCLATQRO 520
508 EIESTNSREIRKIRIETHLOOQLOESQOMGRLLPEKOIISDOLKQOYO-ONSUHRDSLT 566
521 ELEVLNDQOCDLEIWEIKOLOLEOYOKRLIYLPKOKLNERIKANNQFSNTPPSGVSL 580
567 LKRALEAKELARQOLREQLEVERETRSKLOEIDVFNNO----- 605
581 HKSGLEKELE-ORIKEDLDALKEKTSKLSKSEMSFNNOQLCGMMDSVLQCLLSL 639
606 -----LKELEIHSKQOLOKRSIEAARLKQEOERKSLLEKQ-----EDAORRQER 655
640 NNLFLLKELLETAYNTQOLALEOLYKIRDKLKEIRGRLELMOKKLEDEAARAKQK 699
656 DKOMLEHVOEOPRRPRKHEEDRLKREDSYRKEAEERAKPEMODQOSRLFHQEPK 715
700 ENLMKENIRKEE-----EKOKRIQEKTEQKIQEERKAKEKQKDKD----- 743
716 LATQAPWSTTEKGLTISAQESV-----VYVYRALYPFESRSHEITIQGDIVWDES 770

```

```

Db 744 -----TAKAEKKRETAVALVNYRALYFPAARNHDEMSFNSGDIIOVDEK 788
Oy 771 OTGRPGMLGELKGTGKTGMPPANYAEKIP--ENEVTPRAKPVDTLTSAPAKLALRETPAP 828
Db 789 TVGEHGMVYSGFQNGFCMFPONYEKKMPSSENE-----KAVSKKML----- 830
Oy 829 LPVTSSEBSTPNMNADESSSTWPSSENEKRETDNDMTAAQPSLTVPSAQOLRORSAPFT 888
Db 831 LPPTVLSATS-----TSEEPSSNQPASVTDYQN-VSFSMLTVNTSMQ--KKSAPTR 880
Oy 889 ATATGSSSPVLGEGKEVEGLQAQALYPMRAKDNHNFKPSDVIYTLVEQDMMWPGEVQ 948
Db 881 TVSPG-SVSPHGGQVAVENLKAQALCSWRAKDNHNFKPSDVIYTLVEQDMMWPGEVH 939
Oy 949 GQKGFPPKSYVKLISGPRKSTSDTGPTEPSAPSLKVAS--PAKPAIDGEEFIAMTY 1006
Db 940 GGRGMPKSYVKLIPGSEV-----REPEPALVAAVKKFTSAAYSGEEYIALY 991
Oy 1007 ESSGEGDITFQOGDIVYVTKKDGWMVTGYDGSGVFPSPNYVRLKDESGGTAGTGSIG 1066
Db 992 SSVEPGDITFTEGEBELITVQKDGEMWMTSGIDRSGLFPSPNYVPRKDESFGSASKSGASN 1051
Oy 1067 KKEPAQVIAVYATGAPOLTLAPGOLILIRKKNPGGMEGELQARGRKQIGWFPANYV 1126
Db 1052 KKEPAQVIAVYATGAPOLTLAPGOLILIRKKNPGGMEGELQARGRKQIGWFPANYV 1111
Oy 1127 KLSPTGSKITPTLPTKAVGPAVCQVIGMYDYTAONDELAFESKQIINVANKEDPDW 1186
Db 1112 KLGPSSEKATPAPF-----VCQVIAVYDVAANNEDELFSKQIINVANKEDPDW 1164
Oy 1187 KGEVSGGVGLFPSPNYVKLTMDPSQ 1213
Db 1165 QGEINGVTGLFPSPNYVKMTTDSQ 1191

RESULT 6
ID IN2 MOUSE STANDARD; PRT: 1658 AA.
AC Q9ZOR6; Q9ZOR5;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Intersectin 2 (SH3 domain-containing protein 1B) (EH and SH3 domains
  protein 2) (EH domain and SH3 domain regulator of endocytosis 2).
GN ITSN2 OR SH3D1B OR ESE2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX MEDLINE=99164083; PubMed=1064583;
RA Sengar A.S., Wang W., Bishay J., Cohen S., Egan S.E.;
RT "The EH and SH3 domain ESE proteins regulate endocytosis by linking to
  dynamin and Eps15."
RL EMBL J. 18:1159-1171 (1999).
CC - FUNCTION: Adapter protein that may provide indirect link between
  the endocytic membrane traffic and the actin assembly machinery.
  May regulate the formation of clathrin-coated vesicles.
CC - SUBUNIT: Belongs to a complex that may contain multimers of ITSN1,
  ITSN2 and Eps15, and different partners according to the step in
  the endocytic process.
CC - SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC - ALTERNATIVE PRODUCTS: 2 isoforms; 1/Ese2L/long (shown here) and
  2/Ese2/short; are produced by alternative splicing.
CC - TISSUE SPECIFICITY: Widely expressed in adult tissues.
CC - DEVELOPMENTAL STAGE: Widely distributed throughout the adult
  forebrain. Prominent expression was observed in the neocortex, the
  piriform cortex, the pyramidal cell layers of hippocampus, the
  dentate gyrus, in several nuclei of the thalamus and hypothalamus
  and in the amygdala.
CC - SIMILARITY: CONTAINS 2 C2 DOMAINS.

```

CC - SIMILARITY: CONTAINS 1 DBL-HOMOLOGY (DH) DOMAIN.  
 CC - SIMILARITY: CONTAINS 1 EF-HAND CALCIUM-BINDING DOMAIN.  
 CC - SIMILARITY: CONTAINS 1 EF-HAND CALCIUM-BINDING DOMAIN.  
 CC - SIMILARITY: CONTAINS 5 SH3 DOMAINS.  
 CC - SIMILARITY: CONTAINS 2 EH DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sid.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF132479; AAD19747.1; -  
 DR EMBL: AF132480; AAD19748.1; -  
 DR HSP: P29355; 1SEM.  
 DR MGI: 1138049; Sh3d1B.  
 DR InterPro: IPR000008; C2.  
 DR InterPro: IPR002048; EF-hand.  
 DR InterPro: IPR000261; EFS15\_repeat.  
 DR InterPro: IPR001849; PH.  
 DR InterPro: IPR000219; RhGEF.  
 DR InterPro: IPR001452; SH3.  
 DR Pfam: PF00018; SH3; 5.  
 DR Pfam: PF00036; etHand; 2.  
 DR Pfam: PF00168; C2; 1.  
 DR Pfam: PF00169; PH; 1.  
 DR Pfam: PF00621; RhGEF; 1.  
 DR PRINTS: PR00452; SH3DOMAIN.  
 DR SMART: SM00239; C2; 1.  
 DR SMART: SM00054; EFh; 2.  
 DR SMART: SM00027; EH; 2.  
 DR SMART: SM00233; PH; 1.  
 DR SMART: SM00325; RhGEF; 1.  
 DR SMART: SM00326; SH3; 5.  
 DR PROSITE: PS50004; C2\_DOMAIN\_2; 1.  
 DR PROSITE: PS50010; DH\_2; 1.  
 DR PROSITE: PS00018; EF\_HAND; 1.  
 DR PROSITE: PS50031; EH; 2.  
 DR PROSITE: PS50003; PH\_DOMAIN; 1.  
 DR PROSITE: PS50002; SH3; 5.  
 DR Endocytosis: SH3 domain; Repeat; Coiled coil; Calcium-binding;  
 DR Alternative splicing.  
 FT DOMAIN 21 109 EH 1.  
 FT CA BIND 66 78 EF-HAND (POTENTIAL).  
 FT DOMAIN 244 333 EH 2.  
 FT DOMAIN 364 716 COILED COIL (POTENTIAL).  
 FT DOMAIN 717 778 SH3 1.  
 FT DOMAIN 851 909 SH3 2.  
 FT DOMAIN 941 999 SH3 3.  
 FT DOMAIN 1013 1077 SH3 4.  
 FT DOMAIN 1087 1146 SH3 5.  
 FT DOMAIN 1169 1356 DH.  
 FT DOMAIN 1395 1505 PH.  
 FT DOMAIN 1517 1613 C2 DOMAIN.  
 FT VARSPDIC 1187 1197 DPEQVIEVFQ -> GLQFEEQKTL (IN ISOFORM  
 FT 2).  
 FT VARSPDIC 1198 1658 MISSING (IN ISOFORM 2).  
 FT CONFLICT 451 452 KQ -> NT (IN REF. 1; AAD19748).  
 FT CONFLICT 461 1161 D -> G (IN REF. 1; AAD19748).  
 FT SEQUENCE 1161 1161  
 SQ SEQUENCE 1588 AA; 188776 MW; 7050EFC5F7983A5 CRC64;  
 Query Match 48.6%; Score 3049.5; DB 1; Length 1658;  
 Best Local Similarity 52.4%; Pred. No. 3.5e-112;  
 Matches 658; Conservative 177; Mismatches 275; Indels 145; Gaps 32;

DB 61 IVALDLNKGDMDOEFSIAMKLIKLOQOLPVVLPIMKOPMPSPLISA-RFAMG 119  
 QY 119 GIASNP-----PLTAVAPV---PMG-----SIPVNGSPPLSVPPAAPPANAP 164  
 DB 120 ---SNPNLSIHOPLPVAPVPLATPLSSATSGTSPILMPALPVSVSTSLP---NGTAS 173  
 QY 165 VIQPLPAVAPPAATPKSSPS--RSGPSQNLTKQAQS-PVASAP----- 211  
 DB 174 LIQPL-SIPYSSSTLPHASSYSLVMGCG---GASIQAGSLIDGSSSTSTASLSCN 229  
 QY 212 ----AAEWAPQSSRLKYROLFNSHDKTMSGHLTGPOARTLMQSSIPQALASIMNLS 266  
 DB 230 SPKGTSEWAVPQPSRLKYRKFNSLDMKMSGYLSGFQARALLQSNLSQTLATITLA 289  
 QY 267 DIDQGLTAEBFILLAMLIIVAMSGOPLPVLPPEYIPSEFRVRSSSGMSVSSSVD 326  
 DB 290 DIDGGQLKAEBFILLAMLTMAKAGOPPLTLPPELVPSFR---GGKOV---DSVN 341  
 QY 327 QRLPEPSEDEQOEKKLPVTFEDKKENFERGSVELEKRRQALLQORKEORLAQLE 386  
 DB 342 GLTPSYQKTQ-BEEPQKTLPTFEDKKRANFERGMELKRYLMQOQREARKQKE 400  
 QY 387 RAEQKEREEROEAKQOLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLE 446  
 DB 401 KEWERKQRELOEOWKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLE 460  
 QY 447 EWERRROELLNORKEQEGVVLKARKTLEFLEALNDKQHLEGLOIRRLATQK 506  
 DB 461 EWERRROELLNORKEQEGVVLKARKTLEFLEALNDKQHLEGLOIRRLATQK 520  
 QY 507 QEISTNKSRELRIAEITLHLOOQLOESQOMLGRILPEKQILSDLOKQVQNSLRDLSLT 566  
 DB 521 TELEVLDMQCOLEIMEIYQLOQELKEYONKLIYVPEKQILNEIKMQLSNTDSSGL 580  
 QY 567 LKRALFAKELARQOLREOLDEVERETRSKLOEIVFNQNLKEIRHSKQOLKORSLEA 626  
 DB 581 LHKSSKEKEELCORLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLE 640  
 QY 627 ARLKQKEQERSLE-LEKQ---EDAQRVQERDKOMLEHQOQEOQRRPRRHEEDLTK 682  
 DB 641 KRDKLKEIRKRLQLOKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLEKQLE 694  
 QY 683 EDSVAKKEAEERAKPEMDQKQSLFPHQOEPAKLATOAPWSTTEKGLTISAQESVYVY 742  
 DB 695 EKSQKQTEER---KAKQSE-----TASA-----LVN 721  
 QY 743 YRALYPPESRSHDEITQPGDIIVWVDESQGEPEGWLGELKGTGMPYANAEKIPENEV 802  
 DB 722 YRALYPPFARNHDEMSFSSGDIIOVDEKTVGEPGLYGSFGKFGMPVYVEKVLSE- 780  
 QY 803 PTPAPFVDTLSAPAPKALRETPAPLPVTSSESTTPNNMADSSSTWPPSSNKPETDN 862  
 DB 781 -----KALSPKALLPPTVLSLATSTS--SOPASVATVHNV--SFSNLTVNT-- 824  
 QY 863 WDTWAQPSLITVPAGOLQRSATFTPATGSSSPVLAGQEGVGAQALYPMRAKD 922  
 DB 825 --TW-----QKSAFTRTVSPG-SVSPHGGQGAQVBNLKAQALCSWTAKKE 867  
 QY 923 NNLNFKSDVITVLEQODMMWFGVQOKGMPKSYVYKLSG-EVRRS--TSIDTGPES 979  
 DB 868 NNLNFKSDVITVLEQODMMWFGVQOKGMPKSYVYKLSG-EVRRS--TSIDTGPES 927  
 QY 980 PASLKRVASPAKPAIP-GEEFIAMTYESESQODLFFQOQDVIVTKKQDMWTGVGD 1038  
 DB 928 PFS--TAYPTSTRVAPVGEDVIALYSSVSEPDLLFTGEELLVQKQGEWWTGSIGE 984  
 QY 1039 KSGVPSNYVYRLKDSGSGTSGTKGSLGKPELTAQVYASAAATGPOLTAPOGLIIRK 1098  
 DB 985 RTGIFPSYVYVAPKQENPNQANSKSGASNKKPELTAQVYASAAATGPOLTAPOGLIIRK 1044  
 QY 1099 KNPGKMEGELQARQKQ 1158

Db 1045 KNTSGWMOGELQARGKQKQWFPASHVKLLGPSSERTMPT-----FHAVQVIAMWD 1097  
 QY 1159 YTAONDELAFSGQIINLINKEDPDMWKGVEVSGQVGLFPSPNVYKLTMDPSQ 1213  
 Db 1098 YMANDELINFSQGLINWANKDDPPDMOGETNGLTGLFPSPNVYKMTTDSDSQ 1152

RESULT 7  
 EPI5 HUMAN  
 ID EPI5 HUMAN STANDARD: PRT; 896 AA.

AC P42566;  
 DT 01-NOV-1995 (rel. 32, Created)  
 DT 01-NOV-1995 (rel. 32, Last sequence update)  
 DT 15-JUN-2002 (rel. 41, Last annotation update)  
 DE Epidermal growth factor receptor substrate 15 (Protein Eps15) (AF-1P protein).  
 GN Eps15 OR APTP.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Melanoma;  
 RX MEDLINE=94239734; PubMed=8183552;  
 RA Wong W.T., Kraus M.H., Carlomagno F., Zelano A., Druck T.,  
 RA Croce C.M., Huebner K., di Fiore P.P.;  
 RT "The human eps15 gene, encoding a tyrosine kinase substrate, is  
 RT conserved in evolution and maps to 1p31-p32.";  
 RL Oncogene 9:1591-1597(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94181254; PubMed=8134107;  
 RA Bernard O.A., Mauchaulte M., Mecucci C., van den Berghe H.,  
 RA Berger R.;  
 RT "A novel gene, AF-1p, fused to HRX in t(1;11)(p32;q23), is not  
 RT related to AF-4, AF-9 nor ENL.";  
 RL Oncogene 9:1039-1045(1994).  
 CC - FUNCTION: INVOLVED IN CELL GROWTH REGULATION. MAY PLAY A ROLE IN  
 CC - SIGNAL TRANSDUCTION AND MITOGENICITY.  
 CC - SUBUNIT: Interacts with Stonin 2.  
 CC - SUBCELLULAR LOCATION: Cytoplasmic (Probable).  
 CC - TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED.  
 CC - PTM: PHOSPHORYLATED ON TYROSINE BY EGFR.  
 CC - DISEASE: INVOLVED IN A T(1;11)(P32;Q23) CHROMOSOMAL TRANSLOCATION  
 CC IN ACUTE LEUKEMIAS CAUSING FUSION TO THE TRITHOXAL (MLL OR HRX)  
 CC GENE PRODUCT CONTAINS DNA-BINDING MOTIFS RESULTING IN A  
 CC ROUGE ACTIVATOR PROTEIN.  
 CC - SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.  
 CC - SIMILARITY: CONTAINS 3 EF-HAND CALCIUM-BINDING DOMAINS.  
 CC - SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
 CC EMBL; U07707; AAAS2101.1; -;  
 CC EMBL; Z29064; CA82305.1; -;  
 CC HSSP; P02649; I824.  
 CC GENE; HGNC:3419; EPS15.  
 CC MIM; 600051; -;  
 CC InterPro; IPR002048; EF-hand.  
 CC InterPro; IPR000261; Eps15\_repeat.  
 CC InterPro; IPR003903; UTM.  
 CC Pfam; PF00036; efhand; 4.  
 CC SMART; SM00054; EFn; 3.  
 CC SMART; SM00027; EH; 3.  
 CC PROSITE; PS00018; EF\_HAND; 2.  
 CC PROSITE; PS50031; EH\_3.  
 KW Phosphorylation; Calcium-binding; Repeat; Chromosomal translocation;

KW PROTO-oncogene.  
 FT DOMAIN 15 104 EH 1.  
 FT DOMAIN 128 216 EH 2.  
 FT CA BIND 173 184 EF-HAND 1 (POTENTIAL).  
 FT DOMAIN 224 314 EH 3.  
 FT CA BIND 236 247 EF-HAND 2 (POTENTIAL).  
 FT DOMAIN 599 601 15 X 3 AA REPEATS OF D-P-F.  
 FT REPEAT 623 625 1.  
 FT REPEAT 629 631 2.  
 FT REPEAT 634 636 3.  
 FT REPEAT 640 642 4.  
 FT REPEAT 645 647 5.  
 FT REPEAT 651 653 6.  
 FT REPEAT 654 656 7.  
 FT REPEAT 664 666 8.  
 FT REPEAT 672 674 9.  
 FT REPEAT 692 694 10.  
 FT REPEAT 709 711 11.  
 FT REPEAT 737 739 12.  
 FT REPEAT 798 800 13.  
 FT REPEAT 804 806 14.  
 FT REPEAT 825 827 15.  
 FT DOMAIN 768 850  
 FT MOD\_RES 132 132  
 FT PRO-RICH.  
 FT PHOSPHORYLATION (BY TYR-KINASES)  
 FT (POTENTIAL).  
 FT M -> I (IN REF. 2).  
 FT CONFLICT 822 822  
 FT SEQUENCE 896 AA; 98673 MW; A1B9FB04A07FABEB CRC64;

Query Match 7.8%; Score 486.5; DB 1; Length 896;  
 Best Local Similarity 20.5%; Pred. No. 17e-12;  
 Matches 198; Conservative 150; Mismatches 273; Indels 345; Gaps 29;

QY 15 WAITVEERAKHDDQPLSKPIAGFITDQARNFFQSGLPQVLAQIWMALDMMNDGRMD 74  
 Db 122 WAKVEDAKKDAFDLSIPVNGFLSGDKVKXPVLNKGKLPVDILGRWELSDIDHDGMLD 181  
 QY 75 QVESIANKLIKIKLQGYQVLPSTLPVWKOQPAVIASSAPAGIGIASMPPLTAVAPVM 134  
 Db 182 RDEFVAMFLVYCALE  
 QY 135 GSIPVGNSSPLVSVPPAVPPLANGAPVIGLPAPAPAPATWPKSSFSRSGPSQL 194  
 Db 204  
 QY 195 NTKLQKASPDVAPAPPAEWAPOSSRLKYROLFNSHDKTMSGHLTGPQARTIMQSS 254  
 Db 216 KTV-----WVSPAEKAKYDEIFLKTDMDFVSGLEVRLEIFLTGL 257  
 QY 255 POQLASTIWNLSDDIDQKLTAEFFILAMHLDVAM-SGQPLPVLPPEYTPSFRVRYS 313  
 Db 258 PFTLLAHIMSLCDTRKCGKLSKDFALFHLISQKLKIGIDPPVLPBEMIPPS--DRA 314  
 QY 314 GSGMSVSSSVQRLPREPESSEDOQPEKKLPYTFEDKKENFERGSVELEKRRQALIE 373  
 Db 315 SIQKNITGSSPV-----ADFSAIK-----ELDTLNNEIVD 344  
 QY 374 QQRKEQRIQAQERAEQRRKEREROEPAKQOLELEKQLEQRELEQREERKEIERR 433  
 Db 345 LQREK-----NNVEQDILKEKE-----DTIKQRTSEVQ--DIQDEVQRE 380  
 QY 434 EAKRELEQRQRLWERNRROELINQRNKEQGTVALKARKKLEFELEALNDRKHQLEG 493  
 Db 381 NTNLOKLQKQKO-----QVQELDE-----LDEQKQOLEE 410  
 QY 494 KLQDIRCLATQROEISTNKSRELIAETHLQOQLOESQOMIGRLIPKQILSDQKQ 553  
 Db 411 QLKVRKKCAEBAQIISUK-----AELTSQESQISTYEELAK-----AREELSR 456  
 QY 554 VQONSIRHDSLLTKRALEAKELAQOURLREQUDVEVERETRS-----KLQIDVFNNO 607  
 Db 457 LQOETAE-----LEESVESGAQLEPLQOHLQDSQDEISNMQKMLMKQLENHNSQ 510  
 QY 608 ELREHS-----KOOLQKORSLAEARLQKQBERKSIEL----- 641

DR	Pfam; PF00036; ehand; 4.
DR	SMART; SM00054; EH; 3.
DR	SMART; SM00027; EH; 3.
DR	PROSITE; PS00018; EF_HAND; 2.
DR	PROSITE; PS50031; EH; 3.
KM	Phosphorylation; Calcium-binding; Repeat.
FT	DOMAIN
FT	15
FT	128
FT	216
FT	CA_BIND
FT	173
FT	184
FT	DOMAIN
FT	224
FT	314
FT	CA_BIND
FT	236
FT	247
FT	599
FT	829
FT	DOMAIN
FT	599
FT	REPEAT
FT	601
FT	REPEAT
FT	623
FT	625
FT	REPEAT
FT	629
FT	631
FT	REPEAT
FT	634
FT	REPEAT
FT	636
FT	REPEAT
FT	640
FT	642
FT	REPEAT
FT	645
FT	647
FT	REPEAT
FT	651
FT	653
FT	REPEAT
FT	665
FT	667
FT	REPEAT
FT	673
FT	REPEAT
FT	675
FT	REPEAT
FT	693
FT	695
FT	REPEAT
FT	711
FT	713
FT	REPEAT
FT	806
FT	808
FT	REPEAT
FT	827
FT	829
MOD_RES	132
	132
	PHOSPHORYLATION (BY TYR-KINASES)

SO SEQUENCE 897 AA; 98470 MW; 08A0C0D423F873C2 CRC64;

Query Match	7.8%;	Score 486;	DB 1;	Length 897;
Best Local Similarity	21.3%;	Pred. No. 1.8e-12;		

[illegible]

DD 122 MAYA3EDNANI DAI F D3B3F V D0F D30DRAVKF V EUN3ALUF VEI D0KA VME303I DHD0NLUD 101

QY 75 QVEFSIAMKLIKLGQYQLPSTLPVMMQQQPVAISSAPAFGGIGIASMPLTAVAPVPM 134

Db 182 RDEFAMFLVCALE-----KEVPM 203

On 135 GSTPVGMSPTVSSVBPRAVPBT.ANGABVTOBT.DAFANRAATWPKSSSFSPSGPCST. 194

$\frac{2}{3}$

\_\_\_\_\_

\_\_\_\_\_

• •

• •

\_\_\_\_\_

• •

• •

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

• •

• •

• •

\_\_\_\_\_

• •

\_\_\_\_\_

• •

• •

• •

\_\_\_\_\_

[illegible]

255 PQAULAS1WNLSD1DUGK1AEEF1LAMIHL1DVAM-SGUELPVLPPEI1PPSHKRVK 313

Db 258 PSALLAHIWSLCDTKGCCGLSKDQFALAFHLINQKLIKIDPHSLTPEMIPPS---DRS 314

QY 314 GSGMSVISSSSVDORLPEEPSSSEDEOPEKLPVTFEDKKRENFERSGVELERKRALLE 373

315 SICRNTGSSPV.....ADECATY.....ET.DTI.NNETYD 344

[illegible][illegible]

100  
 90  
 80  
 70  
 60  
 50  
 40  
 30  
 20  
 10  
 0

DB 381 SINLQKQAQKQ-----QVQELLE-----LDEQKQLE 410

494 KLQDRCRLATQREIESTNKSRELRIAEITHLQQLQESQQLGRLLPEKQILSDQLKQ 553

Db 411 QLOEVRKKCAEEAQLISSLK-----AETSQESQISSYEEELK-----AREELSR 456

554 VONST.HPDSI.ITI.KPAB.FAKET.APOOT.PEOT.DEVEPE ---TPSKT.OETDVENNOK.E 608

[illegible]





FT KKAHMDPILMTFREFLNQVOMKREKKGNOFLQVSH  
 FT QFHLOVFLHPPHEDLICFL (IN REF. 1).  
 SQ SEQUENCE 1480 AA; 160267 MW; F3518495FF759553 CRC64;  
 Query Match 6.5%; Score 410.5; DB 1; Length 1480;  
 Best Local Similarity 20.5%; Pred. No. 2.6e+09;  
 Matches 261; Conservative 174; Mismatches 444; Indels 395; Gaps 47;  
 17 ITVEERAKHDQOFLSLKPI-AGFTITGDQARNFPOGLPQVLAQIMWADNNDGRMDQ 75  
 266 ITAQDQAKFETLFRSIVTNGSNTVSGANCRLIMRSGLPESQALRIWLTCDTSKAGELLF 325  
 76 VEFSTAMKLIKIKLQCYQULPSTLPVPMKO----- 104  
 326 PEFALAMHLINDVLODITLPYELDSKTKNEVSSFIDAINLSIANDSSANDAKPTPDEF 385  
 105 -----OPVAISSAPAFGI-----GGIASM----- 123  
 386 ITAGVONLOPOTGYWPTQTSFGIPLQSOITGGGVASALNPSTGEMAPPTFMKSMNTGTP 445  
 124 ---PPLTAVAPVPM-----GSIP-VVGMSPPLVSVPPAAV-----PPLAN 160  
 446 GLNPQITGAPASMGQININGNALQPTTGMPQTTGMPQTTGMPQTSFGVNLGPOLTG 505  
 161 GAPPVYQP-----LPFAHAPATWPKSSFSRSGPQOL----- 194  
 506 GA---LQSOYTGAGYGVWPDQSGSPASM--PNLSFNQOGLQSLTGLQPTGFLPSPNFS 560  
 195 -----NTKLQAGSFDAVASAPPAEMWAVPOSRLKYQOLFNSHKMTMSGHLT 241  
 561 ATMPPLTAQTTGGNNETLYKSNFNNNLINDSSODKISTEKSLEFYKIFETPTONKGLLD 620  
 242 GPOARTILMOSSLPOQALASINWLSIDIDQGLTAEFFILAMHLIDVAMSGOPLPVLP 301  
 621 SPTAVETFRKSGINRADLEQINWLCIDINNTGQINKEFALGHVLYGKLGKRGIPNVLP 680  
 302 EYTP-----PSF-----RVRVSG-- 314  
 681 SLIPSTKLIDLKNQKLTPTTTEKPSFGKIDALSYKNNDDVLPNYRNRKRYVASKN 740  
 315 ---SGMSVSSSVDRRLPEPSSSEDOQPEKKLPVTFEDK---KREPFEGSVLEKR 367  
 741 EEOSSSSSSASAVNH-----SSSTLTQTDIDISVDKTVKKTAKPKYAGFSR---EINIK 791  
 368 ROALLEOQKEOER-----LAOLERAEOEKEREOROQEAQRQLE 407  
 792 NIASLENEIKNISNPENCYDSSIPSDLTSRPDAIIAKPLNLNEISTIDNETNKKIQLY 851  
 408 LEKOLEKORELEOREERREKKEITERREAKRELEOROLEWERNRROELLNQRKEQEGT 467  
 852 RKK---NPSIIIGSGPNCGETENDRKAKSRALLR-----ARMSALTGKSTESEDS 899  
 468 VVLKARKKLTLEF-LEALNDKHKQLEGKQDTRCLATATORQEIESTNKRREIRIETHL 526  
 900 LSNEDQOASAEIKRIQOENGKQOEI---TKDIRSSISDISASLKTMTGSMN---ISN- 951  
 527 OQOLOESQOMLGRLLPEKQILSDOLKQVOONSILHRS----- 563  
 952 -QEFERWEGTIGLEDSVRFL--DDLKSNKNSKVTSSPVPSTPTPVVDRSSSYSGOF 1009  
 564 LTLKRALFAKELARQOLEBQDEVERETRSKLOEIDVFN--NQLKELEIHSKOOL--- 618  
 1010 KYAEERAAALKEQAKKREKMLAKFPKNRRNVQSSRSISSENSRQOQIAGSSNLVLP 1069  
 619 ---OKORSLEAKRLKQEOEKSLLEKKE----- 646  
 1070 RATPFQEEKVEVAQPTQPVQSTQPVQPTQPVQPTQPVQPTQPVQPTQPVQPVQNY 1129  
 647 DAQRAVQERDKQWLEHVQOEOQPRPKPHEDRLKR-----EDSVYKKAEE- 693  
 1130 MAKQESDDEDEDEKRLQELKRLKLLKKADEKRLALARKQIDQACNSESBEELNGKD 1189  
 694 -----RAKP-----EMQDKOSRLFPHQEP--KLATQAPWSTTEKGLPITISA 734

Db 1190 NEGVVNVPOAPAPASAFSONSTNAPRSVHAATVPAACKNSGLPSTTMGNPFRKOA 1249  
 735 QESVKKVYVYALYFFESRSHD-BITTIQPGDIWVWDSQTEPGMGLGELKGTGWPPANY 793  
 1250 SAS-----STFFDRAEMQRIORG-----LDDEE--DDQMSDEBENN--VAN 1292  
 794 AEKIPENEVPTP---AKPYTDLTSAPAPLALRETPAP--LPVTSSEPTTPNNMADS 847  
 1293 DNKVEAKIGHPDHARAPV---TAAPLPSV-----TPVPVAVVPOANTSNEXSSPIPIA 1345  
 848 STWSSSNE-----KPETDWMDTMAQPSLITPSAQGLRORS-AFTPATGSSSPSVLG 901  
 1346 PIPPSVTQEPVPVLAPLPAVDGQEPPIPSAPALATAVQKSSSTPALAGVLPPP--- 1402  
 902 QGEVKEGLQAOALYPMRKAQDNHLMFNK-----SDVITVLQOQMMWFGVEYQOK 951  
 1403 --PPLPQOASTSPPIIAHDYVNGAEGKGAIGAGSDSDDVLSIPSS-----VGTBEBEE 1455  
 952 GWPPKSYVKLISGP 965  
 1456 GAQPVSTAGIPISIP 1469  
 Db 1456 GAQPVSTAGIPISIP 1469  
 RESULT 10  
 YBET YEAST STANDARD; PRT; 1381 AA.  
 AC P34216;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Hypothetical 150.8 kDa protein in SEC17-QCR1 intergenic region.  
 GN YBL047C OR YBL0520 OR YBL0501.  
 OS Saccharomyces cerevisiae (Baker's Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OC NCBI\_TaxId=4932;  
 RN [1]  
 RP SEQUENCE OF 1-961 FROM N.A.  
 RC STRAIN=S288C;  
 RX MEDLINE=95176707; PubMed=7871888;  
 RA de Wergifosse P., Jacques B., Jomiaux J.-L., Purnelle B., Skala J.,  
 RT "The sequence of a 22.4 kb DNA fragment from the left arm of yeast  
 RT chromosome II reveals homologues to bacterial proline synthetase and  
 RT murine alpha-adaptin, as well as a new permease and a DNA-binding  
 RT protein.";  
 RL Yeast 10:1489-1496(1994).  
 RN [2]  
 RP SEQUENCE OF 579-1381 FROM N.A.  
 RC STRAIN=S288C;  
 RX MEDLINE=94205266; PubMed=8154187;  
 RA Scherens B., el Bakoury M., Vierendeels F., Dubois E., Messenguy F.,  
 RT "Sequencing and functional analysis of a 32,560 bp segment on the  
 RT left arm of yeast chromosome II. Identification of 26 open reading  
 RT frames, including the KIP1 and SEC17 genes.";  
 RL Yeast 9:1355-1371(1993).  
 CC -1- SIMILARITY: BELONGS TO THE VDP/USOI/YBL047C FAMILY.  
 CC -1- SIMILARITY: CONTAINS 3 EH DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z35808; CA84867.1; -  
 DR EMBL: X78214; CA55048.1; -  
 DR EMBL: Z23261; CA80797.1; -  
 DR PIR: S45781; S45781.  
 DR SGD: S0000143; YBL047C.

DR InterPro: IPR000261; EPRS15\_repeat.  
 DR InterPro: IPR000449; UBA\_domain.  
 DR Pfam: PF00036; ehand; 2.  
 DR Pfam: PF00627; UBA; 1.  
 DR SMART: SM00054; Efn; 2.  
 DR SMART: SM00027; EH; 3.  
 DR SMART: SM00165; UBA; 1.  
 DR PROSITE: PS50031; EH; 3.  
 DR Hypothetical protein; Repeat.  
 KW DOMAIN 14 113 EH 1.  
 FT DOMAIN 135 227 EH 2.  
 FT DOMAIN 277 366 EH 3.  
 SQ SEQUENCE 1381 AA; 150783 MW; 626FD261DCBA7D99 CRC64;

Query Match 6.4%; Score 403; DB 1; Length 1381;  
 Best local similarity 20.0%; Pred. No. 4.7e-09;  
 Matches 286; Conservative 199; Mismatches 486; Indels 458; Gaps 57;

QY 3 QFPRPF-GGS-----LDVVAITVEERAKHDQFSLKPLAGFITGDQARNFFPSGLP 54  
 DB 109 QNPAPMOSGATGNTNNTDIPALSSNDIAKFSQLPDRTKAGCAQTVAGDKAKDI FLKARLP 168  
 QY 55 QPVAQIATLADNMNDGRMDQVEFSIAMKLKTLQGYOLPSTLPVYMKQOPVAISSAPA 114  
 DB 169 NQTLGELMALCDRDAAGVLDKSEFTIMAYLIQLCMSHHPSMNTPEPAVLPTQ----- 219  
 QY 115 FGIGIASMPPLTAVAPVPMGSIPIVVGMSPPVSVPPAAVPLANGAPVYIOLPAPFAH 174  
 DB 220 -----LWDSI---RIEPPVYVNOFPRTTPLSANST 245  
 QY 175 PATWPKSSSFSGSGSGLNTKLOKAQSPVAPPAEAVAVQSSRLYRQLFNSHDK 234  
 DB 246 GVSSLTHTSTISRLSTAGFSA-----ASDKSLSEFKKQGFDAIFDLDK 230  
 QY 235 TWSGHTLGPQARTILMOSSLFQAOLASTIWNLSIDQDKLTAEFFIAMLIDVANSQP 294  
 DB 291 QHAGSLSSAVLPPFLSSRLNQETLATIWDLADIHNNAAEFTKLEFALAMFLIQKNAGVE 350  
 QY 295 LRPVLPPE-----YIPSPFRVRSQSGMSVSSSS----- 324  
 DB 351 LRPVLPPELLOSPLGLYPPNPPLPQQQSAFOIALPSRAKSPLODMHGVSAVAVNTQPT 410  
 QY 325 VDQRLPE-----EFS-----SEDEQOP 341  
 DB 411 VPQVLPQNSNNGSLNDLLALNPSSSPPTKQTVQNTNNSFYDNNNGQATLQQQOP 470  
 QY 342 EKKLPVTPEDEKKRENERGSGVLEKRRQALLQOKKEQERL----- 382  
 DB 471 QQPPLTHSSSGLKPKFTPTS---NFGQSIIKEPPEEQQLRESSDTFSAQPPVPKHAAS 526  
 QY 383 ----- 382  
 DB 527 SPVKRTASTITLPQVPNFSVSPMPAGATSAATGAAGAAGAAALGASAFSSSNNAPFN 586  
 QY 383 -----AQLEBAEQERKERERQEQEAKRQLELEKQELERQREERKEIER 432  
 DB 587 QDLFADGASQSLNATTEMANLSQVNSLSKQASITTDKSKSRATQELKRYTEMKNSIQI 646  
 QY 433 REAKRELERQOLEMENNRQELLNQNKQEGVVLKARKLTLELELELANKKQOLE 492  
 DB 647 KLNLRSTHDNVKQTEBLEAQVL--QVKNENE-----TLAQQLAVEANHYAAE 694  
 QY 493 GKLDIRGLRATQROEISTYKSRRLIAEI---THLQOOLQESQOQMLGRLLPEKQILS 548  
 DB 695 SKLME---LITDQESQTKNAELKEQITNNSMTASIQSOLNKKQ----- 737  
 QY 549 DQLKQVQONSJHRDSLTLTKRALKEKELARQOLBQLDVEYRETRSKQIQEIDVNNQKE 608  
 DB 738 -QVXQ-----ERSWVDVNSKQLEINQTVANLOKEIDGLG-----EKISVYLTKQKE 783  
 QY 609 LREIHSKQOLQKORSLKARLQKQKEQ---RKSLEL---EKQKEDAPORVQCRDQWLEHV 663  
 DB 784 LND-----YQKTVBEQQAQLOAKIQDLSNKDQTLTREKQLEERNRQIEQENLYHQHV 837

QY 664 QOEEQPRPKRHEEDRLKRENSVAKKEABERAK-----PEMODKOSRLFHPHOEP-- 713  
 DB 838 SKIQE-----MFDLSQRKASFEEKADQELKERNIEYANNVRELSKQNMN-AMGQLEP 890  
 QY 714 -----AKLATOAPWST---TEKGPL---TIS-----AQESKVVYVYALYPFESRSHDEI 757  
 DB 891 AKDIIKASASNTDTTKATKATGRGVNHEDTYSKFVETTVENSNNLVNKKDEEKTERTES 950  
 QY 758 TIQPGDI-VWVDESQTEGPGWLGBELKQKGMFPANYAEKIP-----NEVLP-APV 809  
 DB 951 DVFEDVLPVLQSQSDSEBANTNNGQSGNETANP-NLTETLSDFRFDLNEYGIPRSQSL 1009  
 QY 810 TDLTSAPAPKIALRETPAPLPVTSSEPTTNNADEFSSTWPSSSNKEPETDNMDTWAQ 869  
 DB 1010 TSSVANNAPQ-SVRD-DVELPTELEERTI-NNTANRNDT---GUSHIPGEWEALPAT 1062  
 QY 870 PSALTVPASAGOLRORSAPTPATATGSSSPVLG--QGEKVEGLQAO--ALYWRKAKDN- 923  
 DB 1063 ASTDVL-----NETTEVIEDGSTTKRANSNEDGESVSIQSPKISAQ-P-KAXTINE 1114  
 QY 924 -----HLNFKSDVITYLLEQDMMWFGEEVQGGKWPFKSVYKLSIGP----- 965  
 DB 1115 EEPPIQELHIDESSSSSDDDE---FEDTRE---IPSATVKTLPQYNAQPTSSLEIH 1166  
 QY 966 ---VKSTSIDPTSPESPASIKRVASPAKPAIP--GEFTIAMTYESSQGDILTFQGD 1020  
 DB 1167 TEQVIRKYAPAGTSPSHNNGSKKASTNSILPVKDFDFEFAGL-----EQAABEDNG- 1219  
 QY 1021 VIVTVKDDGMWTVGTGDKSGVPSNVYRLKDE-----GSGPAKGTSGSGKPEI 1071  
 DB 1220 -----ADSESEFENVAVAGSMEQEFETIDHLDLQMAFTGLTSSSNPTI-PRVQ 1272  
 QY 1072 AQVASYAATGEQULLAPGQLILRKKNPGMWEGELQAGKRRQGWFPANYVKLSP 1131  
 DB 1273 QQ-----QSTSPAQV-----SNDEND-ELFA-----GF----- 1295  
 QY 1132 GTSKITPTTEL--PKTAVQPAVCOVIGMYDYTAOND--ELAFSGQGIN 1176  
 DB 1296 GNSKAEPTKVATPSPQOP-----IPUKNDPIVDASTLSKGPVYN 1334

RESULT 11  
 TRHY RABIT  
 AC P37709;  
 ID TRHY RABIT STANDARD; PRT; 1407 AA.  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Trichohyalin.  
 GN THH.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxId=9986;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Pletz M.J., Rogers G.E.;  
 RL Submitted (DEC-1992) to the EMBL/Genbank/DBJ databases.  
 CC -!- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES  
 CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE  
 CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR  
 CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY  
 CC ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER  
 CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN  
 CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN  
 CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL  
 CC DIFFERENTIATION.  
 CC -!- SUBUNIT: HOMODIMER (PROBABLE).  
 CC -!- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS  
 CC THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN  
 CC THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).  
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF

CC THE EPIDERMIS.  
 CC -1- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND  
 CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST  
 CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS  
 CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED  
 CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS  
 CC THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.  
 CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN  
 CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG  
 CC DIFFERENT SPECIES.  
 CC -1- PTM: SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE  
 CC PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.  
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE S-100  
 CC FAMILY.  
 CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z19092; CAA79519.1; -  
 DR PIR: S28589; S28589.  
 DR HSRP: P02633; 41CB.  
 DR InterPro: IPR001751; CaBP S100.  
 DR InterPro: IPR002048; EF-hand.  
 DR Pfam: PF00036; efhand.1.  
 DR Pfam: PF01023; S\_100.1.  
 DR ProDom: PD003407; CaBP\_S100.1.  
 DR PROSITE: PS00018; EF\_HAND; 1.  
 DR PROSITE: PS00303; S100\_CaBP; 1.  
 DR KEGG: K01001; Calcium-binding.  
 FT DOMAIN 1 91 S-100 LIKE.  
 FT CA\_BIND 22 33 EF-HAND 1 (LOW AFFINITY) (POTENTIAL).  
 FT CA\_BIND 62 73 EF-HAND 2 (HIGH AFFINITY) (POTENTIAL).  
 SQ SEQUENCE 1407 AA; 183781 MW; A617D2A159F12B7F CRC64;

Query Match 6.2%; Score 391.5; DB 1; Length 1407;  
 Best Local Similarity 29.7%; Pred. No. 1.3e-08;  
 Matches 132; Conservative 97; Mismatches 140; Indels 75; Gaps 18;

QY 326 DQRLPEPSSSEDEQOEKPLVTFEDKKNFSGSVELEKRRQALLLEOQ----- 375  
 DB 198 EEFEIEEQQLRRRQOEQLKRLREBEQRRRRRQHERALQEBEQLLRQRMREPREQO 257  
 QY 376 -----RKQERLAQLERAFQERKERERQOEAKRQLELEKQERLEQREER 426  
 DB 258 QLRRELEIEIRERQRLQERREERQQLRLEQERREERQRLRELEIRERQRLQEE 317  
 QY 427 RKE-----IERRE-AKREL-----ERQRLQERREERQRLQEE----- 473  
 DB 318 RRRQRLQERREERQRLQERREERQRLQEE-----ERREQLAEVEQARERESSTRM 376  
 QY 474 RKLFELELEALNKKQLEGLQDRCRLATQROEISTKSRLEIAITHTLOOQLQ-- 531  
 DB 377 QRLQLEBAGRGSKVSRPRQEEQSLRQOEHRQ-----RQERERLEEQARQOQWQME 432  
 QY 532 -ESQOMLGR1-----IPEKOILSDQLKOVQO-----NSLHRSLLTKRALEKELARQO 580  
 DB 433 EESERRRORLAPRSLRERQRLRAEERQOEORPREERERQRLQEELEEQQLRE 492  
 QY 581 LRQQLQLE-----VERERSLQLEI-----DVFNNOLEK-----LRLEIS-----KOOLQKRS 623  
 DB 493 RAQQLQLEEDSFQSDRERRRQOEQRPEQCTWRWQLEBAQRRRTLLYAKQGOQLREBE 552  
 QY 624 LEARLQKEQERKSLLEK-----QKEDAQREVRQERDKQW-----LEHVQOEOPRPKPEEDR 679  
 DB 553 LQREKRRQ-ERREHYREELQGREBEKRRRQERQRYRLLELRQEBQRLKRLREBEQ 611  
 QY 680 L---KREDSVRKKEAEERAKPEMQ 700

DB 612 LLOEBERLRQERERKLRBEQ 635

RESULT 12  
 TRHY HUMAN STANDARD; PRT; 1898 AA.  
 AC Q07283;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Trichohyalin.  
 GN THH OR TRHY OR THL.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 CX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93280194; PubMed=7685034;  
 RA Lee S.-C., Kim I.-G., Marekov L.N., O'Keefe E.J., Parry D.A.D.,  
 RA Steinert P.M.;  
 RT "The structure of human trichohyalin. Potential multiple roles as a  
 RT functional EF-hand-like calcium-binding protein, a cornified cell  
 RT envelope precursor, and an intermediate filament-associated (cross-  
 RT linking) protein.";  
 RT J. Biol. Chem. 268:12164-12176(1993).  
 RN [2]  
 RP SEQUENCE OF 1731-1898 FROM N.A., AND CHARACTERIZATION.  
 RX MEDLINE=93315897; PubMed=7686953;  
 RA O'Keefe E.J., Hamilton E.H., Lee S.-C., Steinert P.M.;  
 RT "Trichohyalin: a structural protein of hair, tongue, nail, and  
 RT epidermis.";  
 RL J. Invest. Dermatol. 101:65S-71S(1993).  
 CC -1- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES  
 CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE  
 CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR  
 CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY  
 CC ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER  
 CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN  
 CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN  
 CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL  
 CC DIFFERENTIATION.  
 CC -1- SUBUNIT: MONOMER (PROBABLE).  
 CC -1- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS  
 CC THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN  
 CC THE FLIPLOM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF  
 CC THE EPIDERMIS.  
 CC -1- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND  
 CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST  
 CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS  
 CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED  
 CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS  
 CC THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.  
 CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN  
 CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG  
 CC DIFFERENT SPECIES.  
 CC -1- PTM: SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE  
 CC PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMIDASE.  
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE S-100  
 CC FAMILY.  
 CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: L09190; AAA65582.1; -

```

DR PIR; A45973; A45973.
DR HSP; P02633; 4ICB.
DR Genew; HGNC:11791; THH.
DR MIM; 190370; -.
DR InterPro; IPR001751; CapB_S100.
DR InterPro; IPR002046; EF-hand.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00036; eHand; 1.
DR Pfam; PF01023; S_100; 1.
DR Prodom; PD003407; CapB_S100; 1.
DR Prosite; PS00018; EF_HAND; 1.
DR Prosite; PS00303; S100_CaBP; 1.
DR Keratinization; Repeat; Calcium-binding.
KW Keratinization; Repeat; Calcium-binding.
FT DOMAIN 1 91
FT CA_BIND 22 33
FT CA_BIND 62 73
FT DOMAIN 314 390
FT REPEAT 314 326
FT REPEAT 327 339
FT REPEAT 340 351
FT REPEAT 352 364
FT REPEAT 365 377
FT REPEAT 378 390
FT REPEAT 391 444
FT REPEAT 391 396
FT REPEAT 397 402
FT REPEAT 403 408
FT REPEAT 409 414
FT REPEAT 415 420
FT REPEAT 421 426
FT REPEAT 427 432
FT REPEAT 433 438
FT REPEAT 439 444
FT DOMAIN 444 702
FT DOMAIN 923 952
FT REPEAT 923 952
FT REPEAT 953 982
FT REPEAT 983 1012
FT REPEAT 1013 1042
FT REPEAT 1043 1072
FT REPEAT 1073 1102
FT REPEAT 1103 1132
FT REPEAT 1133 1162
FT DOMAIN 1250 1849
FT CONFLICT 1252 1752
FT CONFLICT 1794 1801
FT CONFLICT 1857 1880
FT CONFLICT 1880 1880
SQ SEQUENCE 1898 AA; 247219 MW; A74B5947FB62E31D CRC64;

Query Match 6.1%; Score 381; DB 1; Length 1898;
Best Local Similarity 27.0%; Pred. No. 4.7e-08;
Matches 136; Conservative 117; Mismatches 154; Indels 96; Gaps 21;

```

```

Db 571 REEKRLEQERREORLKRGEERDQLKREERROQLKREERRLQKREERLEORLKEEVERLE 630
Qy 581 LREOLDE-VEBETRSKLOEIDVFNNOLEKLR--EHSKOOLQKORSL---EAPRLKOK 632
Db 631 QERDERDLKREEPREERHLLKSEBQEEERHQLRREOQERREORLKRKEEERLEOR 690
Qy 633 -----EQERKSEL-ENKQEDQRRVQERDKQWLEHNVQGEOPPRK----- 673
Db 691 LKREHEERREOGLAEEOQERERIKSRIPKQWQLESEADAKOSKYLLEAPQAGRAE 750
Qy 674 PHEEDRLKREDSVRKKEAEERAKPEMDOKOSRLFHPQEPAKLATOAPWSTTEKPLTIS 733
Db 751 PQGEKRRRSESLQWGEERAHQQOQEEBQRDPFTQWQAE-----EKSEGRORLS 803
Qy 734 AQESKVVYVYALYFFESRSHDE 756
Db 804 APPPLREQRERQLRAERQOREQ 826

RESULT 13
TRHY_SHEEP STANDARD; PRT; 1549 AA.
AC P22793;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Trichohyalin.
GN THH.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93260018; PubMed=7684041;
RA Fietz M.J., McLaughlan C.O., Campbell M.T., Rogers G.E.;
RT "Analysis of the sheep trichohyalin gene: potential structural and
RT calcium-binding roles of trichohyalin in the hair follicle.";
RL J. Cell Biol. 121:855-865(1993).
RN [2]
RP SEQUENCE OF 1016-1549 FROM N.A.
RX STRAIN=Metino-Dorset horn X Border Leicester; TISSUE=wool follicles;
RC MEDLINE=90130632; PubMed=2298812;
RA Fietz M.J., Pressland R.B., Rogers G.E.;
RT "The CDNA-deduced amino acid sequence for trichohyalin, a
RT differential marker in the hair follicle, contains a 23 amino acid
RT repeat.";
RL J. Cell Biol. 110:427-436(1990).
CC -!- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
CC ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
CC WITH INVOLUCGIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
CC DIFFERENTIATION.
CC -!- SUBUNIT: HOMODIMER (PROBABLE).
CC -!- ALTERNATIVE PRODUCTS: AT LEAST TWO FORMS MAY BE PRODUCED BY
CC ALTERNATIVE SPLICING OF THE SAME GENE.
CC -!- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
CC THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
CC THE EPITHELIA OF THE TONGUE, HOOF AND RUMEN.
CC -!- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STANDED
CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
CC THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
CC DIFFERENT SPECIES.

```

CC -1- PTM: SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE  
 CC PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMINASE.  
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE S-100  
 CC FAMILY.  
 CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-stb.ch/announce/>  
 CC or send an email to [license@isb-stb.ch](mailto:license@isb-stb.ch)).  
 CC -----  
 DR EMBL; Z18361; CAA79165.1; -  
 DR EMBL; X51695; CAA55992.1; -  
 DR PIR; A34209; A34209.  
 DR PIR; S32633; S32633.  
 DR PIR; A40691; A40691.  
 DR HSSP; P02633; 11G5.  
 DR InterPro; IPR001751; CABP\_S100.  
 DR InterPro; IPR002048; EF-hand.  
 DR Pfam; PF00036; efhand.1.  
 DR Pfam; PF01023; S\_100; 1.  
 DR ProDom; PD003407; CABP\_S100; 1.  
 DR PROSITE; PS00018; EF\_HAND; 1.  
 DR PROSITE; PS00303; S100\_CABP; FALSE NEG.  
 KW Keratinization; Repeat; Calcium-binding.  
 FT DOMAIN 1 91 S-100 LIKE.  
 FT CA\_BIND 22 33 EF-HAND 1 (LOW AFFINITY) (POTENTIAL).  
 FT BIND 62 73 EF-HAND 2 (HIGH AFFINITY) (POTENTIAL).  
 FT DOMAIN 413 448 14 X 28 AA APPROXIMATE TANDEM REPEATS.  
 FT REPEAT 449 476 1-2.  
 FT REPEAT 477 504 1-3.  
 FT REPEAT 505 532 1-4.  
 FT REPEAT 533 560 1-5.  
 FT REPEAT 561 588 1-6.  
 FT REPEAT 589 616 1-7.  
 FT REPEAT 617 644 1-8.  
 FT REPEAT 645 678 1-9.  
 FT REPEAT 679 706 1-10.  
 FT REPEAT 707 742 1-11.  
 FT REPEAT 743 771 1-12.  
 FT REPEAT 772 796 1-13.  
 FT REPEAT 797 832 1-14.  
 FT DOMAIN 938 1507 23 X 23 AA APPROXIMATE TANDEM REPEATS.  
 FT REPEAT 938 961 2-1.  
 FT REPEAT 962 985 2-2.  
 FT REPEAT 986 1021 2-3.  
 FT REPEAT 1022 1044 2-4.  
 FT REPEAT 1045 1067 2-5.  
 FT REPEAT 1068 1090 2-6.  
 FT REPEAT 1091 1121 2-7.  
 FT REPEAT 1122 1144 2-8.  
 FT REPEAT 1145 1167 2-9.  
 FT REPEAT 1168 1197 2-10.  
 FT REPEAT 1198 1227 2-11.  
 FT REPEAT 1228 1250 2-12.  
 FT REPEAT 1251 1273 2-13.  
 FT REPEAT 1274 1296 2-14.  
 FT REPEAT 1297 1319 2-15.  
 FT REPEAT 1320 1342 2-16.  
 FT REPEAT 1343 1368 2-17.  
 FT REPEAT 1369 1391 2-18.  
 FT REPEAT 1392 1416 2-19.  
 FT REPEAT 1417 1439 2-20.  
 FT REPEAT 1440 1461 2-21.  
 FT REPEAT 1462 1484 2-22.  
 FT REPEAT 1485 1507 2-23.  
 FT VARIANT 1145 1197 MISSING (IN SHORT FORM).  
 FT VARIANT 1251 1273 MISSING (IN SHORT FORM).  
 FT CONFLICT 1399 E -> G (IN REF. 2).

SQ SEQUENCE 1549 AA; 201173 MW; E72F89FF1326E54E CRC64;  
 Query Match 6.0%; Score 375.5; DB 1; Length 1549;  
 Best Local Similarity 27.6%; Pred. No. 6,2e-08;  
 Matches 116; Conservative 101; Mismatches 114; Indels 89; Gaps 16;  
 QY 332 EPSSDEQPEPKLPVTFE---DKKENFERGSVELEKRRQALLBQKRQELQALR 387  
 Db 383 KPAQEQVREEDQRLKEKQLQREKRQREKQRYVELQRE--ERLQREEDQREER 440  
 QY 388 AEOEKREERQ-----EQAKQOLELKO-----LKQRELERQREER 426  
 Db 441 EKRREKREKQYLEKVELMEBQQLQREBKRRQREKQYLEKVELBEBQQLQREBK 500  
 QY 427 RKEIRREPAARELEQRLQEMERNRQOELQNKQEGTVLAKRKTLFELALND 486  
 Db 501 RQERREQYLEKVELQEBEQLOREERK-----RQRE-----RQYLE----- 538  
 QY 487 KKHQLEGKLDIRCLATQROELSTNKSRELRIAEITTHQ--QQLQSQQMLGRLLPEK 544  
 Db 539 -KVELQEBEQ-----LQREKREKQREKQYLEKVELQEBEQLOREBKQREK 590  
 QY 545 QILSDQLKQVQNSLHSDSLTLKRLAEKELARQQLREQLDVERETRSKLQIDVFN 604  
 Db 591 QYL-EKVELQEBEQLOREKREKQREKQRYLEKVELQEB--EQVQREK-----RR 641  
 QY 605 QUKELREIHSKQQLQKQSLQEARLKQEKQKSLQLEK-----QKEDQOR 650  
 Db 642 QERREQYLEKQLQEBEQLOREBQLREKREKQREKQRYLEKVELQEBEQLOREBK 701  
 QY 651 RVQERDKQWL--EHVQEBQPRPKPHBEDRLKQSDSVKKAEER--AKPNQDKQSL 706  
 Db 702 RQERREQYLEKQLOREB--RLQREKQLOREDEKREKQRYERKYLEBEEQLOREB 758  
 RESULT 14  
 CALD\_HUMAN STANDARD; PRT; 793 AA.  
 ID CALD\_HUMAN  
 AC Q05682; Q13979; Q13978; Q14741; Q14742;  
 DT 01-NOV-1997 (Ref. 35, Created)  
 DT 01-NOV-1997 (Ref. 35, Last sequence update)  
 DT 15-JUN-2002 (Ref. 41, Last annotation update)  
 DE Caldesmon (CDM).  
 GN CALD1 OR CDM OR CAD.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OK NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 4).  
 RC TISSUE=Lung fibroblast;  
 RX MEDLINE=91358497; PubMed=1885618;  
 RA NOVY R.E., Lin J.L.-C., Lin J.J.-C.;  
 RT "Characterization of cDNA clones encoding a human fibroblast  
 RT caldesmon isoform and analysis of caldesmon expression in normal and  
 RT transformed cells";  
 RL J. Biol. Chem. 266:16917-16924(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 4).  
 RC TISSUE=Aorta;  
 RX MEDLINE=9220999; PubMed=1555769;  
 RA Humphrey M.B., Herrera-Sosa H., Gonzalez G., Lee R., Bryan J.;  
 RT "Cloning of cDNAs encoding human caldesmons";  
 RL Gene 112:197-204(1992).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORMS 2, 3 AND 5).  
 RX MEDLINE=93101679; PubMed=1465449;  
 RA Hayashi K., Yano H., Hashida T., Takeuchi R., Takeda O., Asada K.,  
 RA Takahashi E.-I., Kato I., Sobue K.;  
 RT "Genomic structure of the human caldesmon gene";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:12122-12126(1992).  
 CC -1- FUNCTION: ACTIN- AND MYOSIN-BINDING PROTEIN IMPLICATED IN THE  
 CC REGULATION OF ACTOMYOSIN INTERACTIONS IN SMOOTH MUSCLE AND

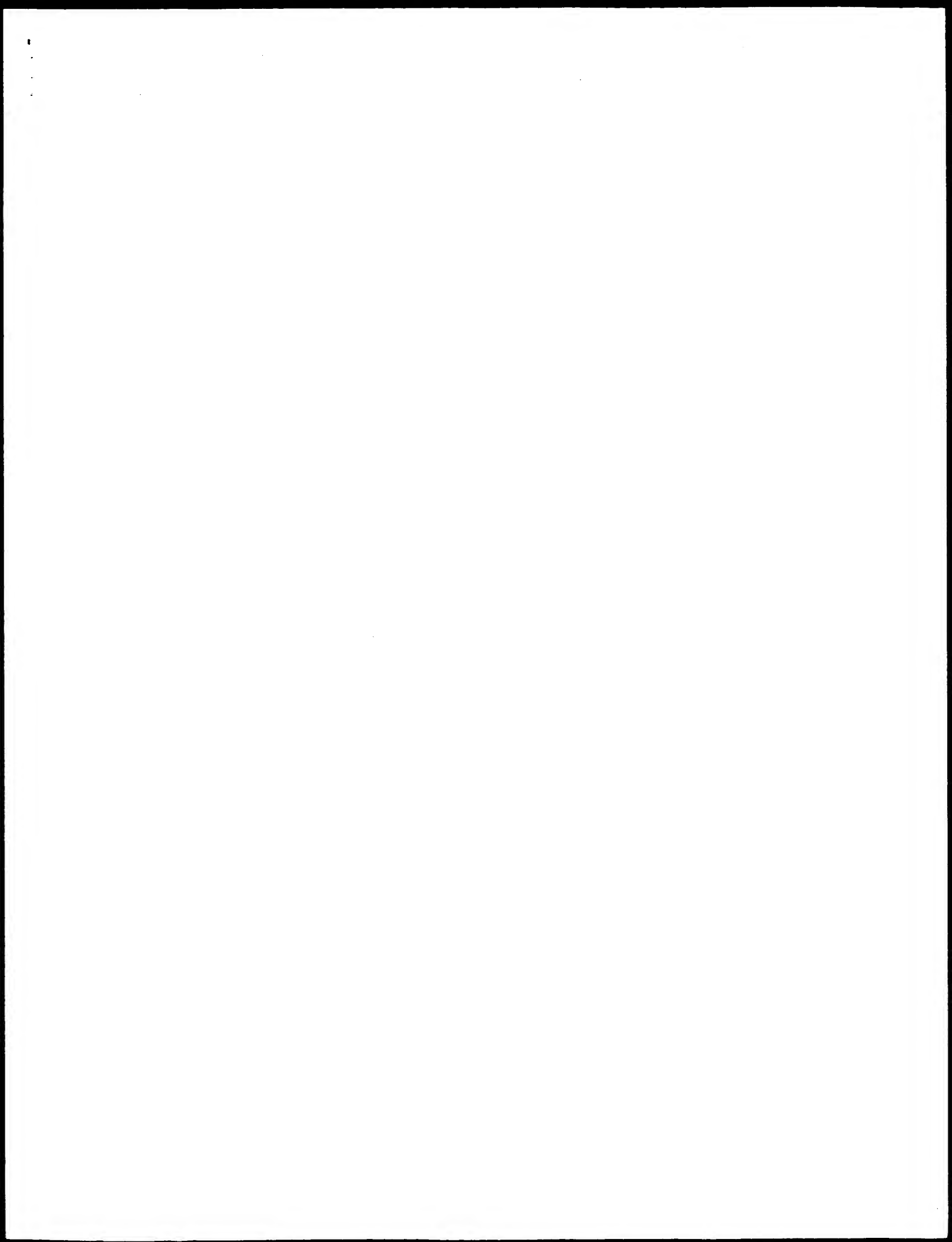
FT	MOD_RES	730	730		PHOSPHORYLATION (BY CLCC2) (BY SIMILARITY)
FT	MOD_RES	753	753		PHOSPHORYLATION (BY CDC2) (BY SIMILARITY)
FT	MOD_RES	759	759		PHOSPHORYLATION (BY CDC2) (BY SIMILARITY)
FT	MOD_RES	789	789		PHOSPHORYLATION (BY CDC2) (BY SIMILARITY)
FT	VARSPLIC	1	24		MDFPERRLRRQREEMLEAE -> MLGGSGSHGRSL AALSO (IN ISOFORM 3 AND ISOFORM 5).
FT	VARSPLIC	208	436		MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT	CONFLICT	530	462		MISSING (IN ISOFORM 4 AND ISOFORM 5).
SQ	SEQUENCE	793 AA;	93250 MW;	2A0D63D16DD6B5F CRG64;	V -> M (IN REF. 1)
<hr/>					
Query Match                      5.3%; Score 330; DB 1; Length 793;					
Best Local Similarity 23.5%; Pred. No. 1.9e-06;					
Matches 150; Conservative 108; Mismatches 209; Indels 170; Gaps 25;					
QY	318	SVTSSSSVDORLPEDPSSEDEQP-----EKKLVTEDEKKNRFEGSVELEKRQ	369		
Db	223	TVMSTLNKGISSEPKOEEREEQGSDIEISHNEK--MEBEDKERAEAKRIEAERER	279		
QY	370	ALLEQRK-----EOERLAQERAEOERKEROEOAPKOLELEKEQERELR	420		
Db	280	IKAEODKIADERARI EAEEKAADOERRRAEREMREEEKRAE-EQORIKEEXNA	338		
QY	421	QREBERKETTEREAARELEROFOLEWERNROELLNORKBOETGVILKANR-KTLFE	479		
Db	339	AEEQRRIKEEKRAE---ERORIKEEKAABERORABEEBEKAVBOQRNKQLER	394		
QY	480	ELEANDKHQLEGKLQDIRCLATOROELS-----TNKSRLRLAIETHLOOOLQS	533		
Db	395	KRRMOETTKIGEKVEQKITEGKWNEKAQEDKIQTAVLKQGEKKTQVQAEREQLQED	454		
QY	534	QQMLGLRIPEKOITLDQLKQVOONSIMHDSLLTKRALAKEIARQQRLQBLDEVERETR	593		
Db	455	KPRP----KKEEIKQE-----KIKQDEKPE---EVKSFMRKKGFTE	490		
QY	594	SKQIEDIVFNQOKELREIHSK-----OOLOKQ-----STE	625		
Db	491	VKSONGFMTHMKIKHTENTFSRPGRASVDTKAAGAOYEAGKRIELELRKRGETSEE	550		
QY	626	AARKQOEORERSLEYE--KQEDAQRAYOERDKOMLEHVQOEOPRPKRPHEEDRLKRB	683		
Db	551	PEFKIQQOQE-AALBELFLKKGBERRRVLEEBEQ--RRQGEADRKLREBEERKLKEE	607		
QY	684	DSVAKKAEERAKPENOD-----KOSRLFHPHOEPAAQLAQAWSTTBKGPLTISOES	737		
Db	608	IERRRAAALKERQKMVEDGLSDDKPFCKCFPPKGSILIERAEF-----LNKSQKS	660		
QY	738	--VVVVVRYRLY-PFERSGHDETITQPGDIVWVDSSQGEFGMGLGKGTGFPANYA	794		
Db	661	SGVASTHQAAIVSKIDSRILEOYT'S-----AIBETSKAKP-----	694		
QY	795	EKIPENEVPTRPAKVTDL-----TSAPAP-----KLAL-----	822		
Db	695	TKPASLDLPAPAGVGNIKSMWEKGVFESSPPAAQTGTPUKETAGUKVGVSRIINEMLTGP	754		
QY	823	--RETAPRLPVYSSSEPTPNNMADPSSRWSPSSMEK	857		
Db	755	DGNKSPAPKP-----SDLRPGDVSSKRNLMWKOSYDK	786		
<hr/>					
RESULT 15					
ID	YAVL_SCHPO	STANDARD;	PRT;	1794 AA.	
AC	Q10172;				
DT	01-OCT-1996 (Rel. 34, Created)				
DR	01-OCT-1996 (Rel. 34, Last sequence update)				
DI	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Hypothetical protein C27F1.0lc in chromosome I.				
GN	SPAC27F1.0lc OR SPAC25G10.09C.				

OS Schizosaccharomyces pombe (fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomyces.  
 RN NCBI\_TaxID=4896;  
 RP [1]  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Soutos J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Church C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jaegels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean K.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymoprez B.,  
 RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Medler H., Wambutt R., Fumelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucac M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Bartell B.G., Nurse P.,  
 RT "The genome sequence of Schizosaccharomyces pombe".  
 RL Nature 415:871-880(2002).  
 CC - SIMILARITY: CONTAINS 2 EH DOMAINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL: Z69368; CA93290.1; -  
 DR EMBL: Z70691; CA94638.1; -  
 DR InterPro: IPR002048; EF-hand.  
 DR InterPro: IPR000261; EFS15\_repeat.  
 DR InterPro: IPR003124; WH2.  
 DR Pfam: PF00036; efhand; 3.  
 DR Pfam: PF02205; WH2; 1.  
 DR SMART: SM00027; EH; 2.  
 DR SMART: SM00246; WH2; 1.  
 DR PROSITE: PS50031; EH; 2.  
 DR Hypothetical protein; Repeat.  
 KW Hypothetical protein; Repeat.  
 FT DOMAIN 281 369 EH 1.  
 FT DOMAIN 821 910 EH 2.  
 SEQUENCE 1794 AA; 193279 MW; 1F042BA5E805E27 CRC64;  
 Query Match 5.2%; Score 327; DB 1; Length 1794;  
 Best Local Similarity 20.9%; Pred. No. 5; 6e-06;  
 Matches 237; Conservative 137; Mismatches 404; Indels 356; Gaps 52;  
 QY 28 QPILSKPIAFITIGDARNFFQSGLPVLAQIMALADN--NNDGRMVERSIMKLI 85  
 DB 676 QRTGMOQPMQRTQMOPQ---MPGMQOPMAQRTGQPMQPMQRTQMOPQ----- 721  
 QY 86 KKLIOGYQLPSTIPVWKOQPAVIASSAPFGIGIASMPPLTIVAVPMGSIPIV----- 140  
 DB 722 ---MPGMQOPMA-QRTGMOQPMQMA-----PRTGMOQPMQMG-PMLQRTG 761  
 QY 141 GMSBPVIVSVPPAAMP-----PLANGAPVIVQPLPAFPAAPATWPKSSSPRSRSGPS 192

DB 762 GMAP-----QPTGPGQWGFINTPLSN--LPGIEALQOQMPMA-----PSG 801  
 QY 133 QUNTLQKQAFDVAAPPAEMWAVPOSSRLKYROLFPSSHDKTSGHLTGQARTILMOS 252  
 DB 802 GINTTFO--QKKDI-----PMALSKKEKRIYDQIFDAMDKEKRTGLGNAVLIEIGQS 852  
 QY 253 SLPOALASIMWLSIDIDQKLTAEELFAMHLIDVAMSGQLPVPVPEVYIPSPFRVR 312  
 DB 853 KLTRELHINWLCDHGKGSLDRDEFVALHLIYKRLNGEVAVALPELIPSTSTNFT 912  
 QY 313 SSGSGSVSSSVQDRLPEEPSSDEQOPEKLLVTEDEKRENFEGSV--ELEKRRQ 369  
 DB 913 --ESLNVQKNLIKNDTSNRKPKFGAENOSKLLK--NSFYDNPSETTEKDATLYRHNDSDAS 968  
 QY 370 ALLEQOREOEHLAEQERAEQERKEREQOEAKROLEQL-----EKORELEQRE 423  
 DB 969 AVSSARRDRKREKIESAPPIINDIDELISLKKRIH-EKSLVNALEDKLLATPAND 1027  
 QY 424 EERRKEIRREAPAKRELEBROROLEWERNROELNORKEQGVTLKARKTLEFELIA 483  
 DB 1028 VQNDSLIYRIKSVQDEINR-----LSTNKSPEVA-----SNMVRLEE 1065  
 QY 484 LNDKQHOLEGKLDIQRCLATQROEISTNKSRLRI--AEIT--HLOQLQESQMLGR 539  
 DB 1066 LSTRVSKM---LSDI-----NEVDHTIASLSLKLFQADRTNSYDQTSPEATOERNR 1114  
 QY 540 LIPEQILSDQKQVQOQSLHRSDILTLKRLAEKELARQOLQREQEVERETSKY--- 596  
 DB 1115 TISSK--LAEKEKQNES-----KALE-----OMKQVYVNIENNIKALPLPS 1155  
 QY 597 -----OEIDVFNQULKEIRHISKOOLQOKOSLEA---ARLK- 630  
 DB 1156 AANDAWLSQWVDESVRVYVKELPVPAAPQTLNPPSVSTVQSKPIESHTTPVKA 1215  
 QY 631 QKEQKSLLELEKQ---KEDAQRRVQERDKOMLEHYQOEQPPPKPHEDRLKREDSV 686  
 DB 1216 TSEPSASNSNEDRAARKAORRMERLALGIRKOKTSPSPA-----PV 1263  
 QY 687 RKKEEERKAKEMOQKQ-----SRLFP-----HOEPKALATQAPW 722  
 DB 1264 NSATSTPVAAATLQOIOGKQASAVSNVPAVSISITPPAVVTQVHPQFTK--QIP- 1319  
 QY 723 STTEKPLTLTSAOESVKKVYVRYALYPESR----- 752  
 DB 1320 TAAVADPSTTSTSFNTATIPQA--PLENQFSKMSLEPPVPAVPTSPKQIIPSSNVHA 1377  
 QY 753 -----SHDEITIQGDIWMDESQTEGPGMLGELGKTKGFPANVAEK----- 796  
 DB 1378 PPPVQPMNAMPVSHNAVAPSPAPERRDSFGSVSSGSNVSIIDETSTMPKASQPTNPG 1437  
 QY 797 IPENEVV-----TPAKPVTD--LTSAPAPKTLARTPAPLVLTSTSEPTSTNNW 843  
 DB 1438 APSNAPQVVPVPAHNAVAPQAPGKMTNAPAPSSA---PAPAPVSQLPAPVAVN-- 1491  
 QY 844 ADFSSTWPSSENKRETDNMTWAQPSLTVPSAGQLRQSAFAFPATATGSSPSPVLGQG 903  
 DB 1492 VVPEVMISVAQOPSS---VAAPATAPSSLTLPSS---QSSF---AHVPSAPAPAPQH 1539  
 QY 904 EKVGLQAOALVPMRAKKNHLNFNKSVDITVLEQQDMWFGVEYQGGKMGFPPKSVKLIS 963  
 DB 1540 PSAALTS-----APADNMPHRRSP-----YAPQD----- 1565  
 QY 964 GPKVRSSTIDT-----GPTD---SPASLKRVASPA-AKPAIP 996  
 DB 1566 -PVQKPAIINNIPATNLGTSQSPSPMGVNNSGSPPLAMAAQPSLAVPAVP 1618

Search completed: March 14, 2003, 12:08:12  
 Job time : 47 secs





GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 14, 2003, 11:19:11 ; Search time 40 Seconds  
(without alignments)  
6248.379 Million cell updates/sec

Title: US-09-674-237A-3  
Perfect score: 6269  
Sequence: 1 MAQFPTPGSLDVMATVE.....VGLFSPSYVKLTDMPSQ 1213

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_21: \*  
1: sp\_archaea: \*  
2: sp\_bacteria: \*  
3: sp\_fungi: \*  
4: sp\_human: \*  
5: sp\_invertebrate: \*  
6: sp\_mammal: \*  
7: sp\_mhc: \*  
8: sp\_organelle: \*  
9: sp\_phase: \*  
10: sp\_plant: \*  
11: sp\_rodent: \*  
12: sp\_virus: \*  
13: sp Vertebrate: \*  
14: sp Unclassified: \*  
15: sp\_rv1rus: \*  
16: sp\_bacteriap: \*  
17: sp\_archaeap: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1735	27.7	1011	5	O61639 drosophila
2	1702	27.1	1097	5	O9VIF7 drosophila
3	1698.5	27.1	1094	5	O61618 drosophila
4	1333	21.3	1097	5	O9U2T9 caenorhabdi
5	794.5	12.7	662	5	O8T068 drosophila
6	536.5	8.6	662	4	O9UBC2 homo sapien
7	528	8.4	1253	5	O8W61 drosophila
8	527	8.4	907	11	O6O902 mus musculu
9	506.5	8.1	751	5	O23418 caenorhabdi
10	506.5	8.1	796	5	O9B1F4 caenorhabdi
11	493.5	7.9	599	11	O91WH8 mus musculu
12	485	7.7	668	5	O9W11 drosophila
13	434.5	6.9	1116	3	O9HGL2 schizosacch
14	405	6.5	944	5	O18137 caenorhabdi
15	405	6.5	1148	5	O952Y7 caenorhabdi
16	405	6.5	1480	5	O18138 caenorhabdi

17	402.5	6.4	1003	12	O91IX9 kaposi's sa
18	397	6.3	1089	12	O40947 kaposi's sa
19	386	6.2	1036	12	O9DUM3 kaposi's sa
20	386	6.2	2376	5	O966V1 drosophila
21	383.5	6.1	1909	5	O25893 plasmodium
22	382	6.1	1162	12	O98148 kaposi's sa
23	377.5	6.0	1129	12	O9QR71 kaposi's sa
24	375	6.0	1151	4	O9BX65 kaposi's sa
25	367	5.9	976	12	O9DUN0 kaposi's sa
26	363.5	5.8	2701	4	O9Y520 homo sapien
27	356	5.7	1743	5	O966V0 drosophila
28	350.5	5.6	1108	5	O9ND10 babesia big
29	350	5.6	1750	3	O8X0H2 neosporea
30	344.5	5.5	1004	4	O9UNH7 homo sapien
31	344	5.5	330	11	O35146 mus musculu
32	344	5.5	420	11	O08635 mus musculu
33	342	5.5	2722	5	O19135 caenorhabdi
34	341.5	5.4	1287	5	O9NGS5 drosophila
35	340.5	5.4	800	5	O9Y102 drosophila
36	336.5	5.4	1292	4	O9BX66 homo sapien
37	332.5	5.3	1639	5	O9VNF2 drosophila
38	331.5	5.3	1069	4	O75137 homo sapien
39	331.5	5.3	1100	4	O9NV82 homo sapien
40	330	5.3	964	13	O42489 gallus gall
41	329.5	5.3	585	11	O9JRK1 rattus norv
42	329.5	5.3	665	11	O925Q9 rattus norv
43	328.5	5.2	554	5	O9VPS3 drosophila
44	326.5	5.2	665	11	O8R551 mus musculu
45	324.5	5.2	1181	10	O9X116 arabidopsis

## ALIGNMENTS

### RESULT 1

ID O61639 PRELIMINARY; PRT; 1011 AA.  
AC O61639;  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
DE Dyanin associated protein isoform DAP160-2.  
GN DAP160 OR CGI109.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_taxid=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OREGON-R; TISSUE=HEAD;  
RA Roos J, Kelly R.B;  
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: CONTAINS 4 SH3 DOMAINS.  
DR EMBL; AF054612; AAC39139.1; -.  
DR HSSP; P29354; IGR1.  
DR FLYBase; FBgn0023388; DAP160.  
DR InterPro; IPR002048; EF-hand.  
DR InterPro; IPR000261; ERS1\_repeat.  
DR Pfam; PF00036; EFhand; 2.  
DR PRINTS; PR00452; SH3DOMAIN.  
DR ProDom; PD000066; SH3; 4.  
DR SMART; SMO0054; EFh; 2.  
DR SMART; SMO0027; EH; 2.  
DR SMART; SMO0326; SH3; 4.  
DR PROSITE; PS00018; EF\_HAND; UNKNOWN\_2.  
DR PROSITE; PS50002; SH3; 4.  
KW SH3 domain.  
SQ SEQUENCE 1011 AA; 111619 MW; 54B589C92021C91 CRC64;

Query Match 27.7%; Score 1735; DB 5; Length 1011;

Best Local Similarity 33.4%; Pred. No. 1,5e-73;  
Matches 413; Conservative 202; Mismatches 332; Indels 286; Gaps 31;

QY 11 SLVDAWTAVERAKHDOOFLSLKPIAGITTDQARNFFSGLPVLAQIWMALAMND 70  
D 4 AVDMAVAVPRERLKYQEOFRALQPOAGFVTAQAKGFLOQLPPLILGQIMALADTD 63  
QY 71 GRMOQVESIMKLIKILKLOGYQULPSTLPVYKQOQPAVISAAPAFGGIASMPP--LT 127  
D 64 GKMINNESIACKLINLTKRGMDVFKVLPSEL-----LSL---TGDPVSMTPRSGSTS 113  
QY 128 AVAPVPMGSIENVGMSPELVSVPPAAPPVPLANGAPVYIOPLPAPAPATWPKSSFSR 187  
D 114 SLSPLD-----PLKGI-VPAVAVVVPVAVPVA--VATVISP-PGVSPSGPTPTSTN--- 162  
QY 188 SGPSQNLTKQAKQSPVAVAPPAEVAWVQSSRLKROLFNSHDKTMSGHLTGPAQT 247  
D 163 --PSRHTSISERAPSEISVN---QGEWAQAQKRKTYOVFNANDRTSGYLTGSGARG 217  
QY 248 ILMOSLPQAOLASIMWNSDIDODGKLTAEFFILAMHLDVAMSGLPPLVLPPEYIPPS 307  
D 218 VLVQSKLPQVTLAQITWLTSDIDGRLNCDFFILAMPCEAMAGKIPVTLPEWVPPN 277  
QY 308 FRVRSQSGMSVSSSVQDRLPEPSSD-----EQQPEKLP-VTEPKK 353  
D 278 LRKTSRPG---SVSGVSRPSQSPASRHASVSSQGVVADADPTAGLPQTSFEDKR 333  
QY 354 RENBERGVELEKRRQALEOORKEOERLAJERAEOERKREOEOERAKOLELEKOLE 413  
D 334 KENVYKGAELDRRKRLKEDQORKEERERERERERERERERERERERERERERERER 393  
QY 414 KQELEROREER 473  
D 394 RQREIEMKEQORRELAKAEARKELEKQOQWEOQRIENNAQAQREBERVLAQKXAH 453  
QY 474 RKTLEFEIANDKKHOLEGKQDRIKCLATROEISTMSRRLRAETHLOOQIOES 533  
D 454 NTOINVELISTINEKIKELSORICOTRAVTVKVIDCMRTQRTDSSEMSQLARIKEQ 513  
QY 534 QOMGLRLPER-----QILSDQKOVQNSLHRDLSLTKALEKEKLAQOQLR 582  
D 514 NAKLTQLOEAKWEAKSKASGALGNGNAQEO-----LNAFAHQIILNIOIK 563  
QY 583 EOLDEVERETRSKLOEIDVFNQKLEIRE-----IHSKQOLQOKRSLAARLKQEKERK 637  
D 564 DKVENISKEISKEDIDITNDVQMSLEKALITKEDIDYKREYDQRTSVLELTKYRK 623  
QY 638 SLELEKQKEDAKORVQOERDKWLEHVOOEOPRRKPHHEBRLKREDSYKKEKEERAKP 697  
D 624 NETSVSSAMDT---GSSSAM---EETGTTVTDPY-----AVASNDISALAAP 664  
QY 698 EMODKOSKRLFPHOEPAKLATQAPWSTTEKGLTISAQESVYVYALPFFESRSHDI 757  
D 665 AVD-----LGGPAP-----EG-----FKYQAVVEFNARNAEI 693  
QY 758 TIGPEDIWVDESQTEGPGWLGBELKGTGMFPANVAEKIPENEPVPPAKVTDLTSPA 817  
D 694 TFEVGDITLVLEONAEHGLAGEINGHTGFWPESYVEKLEVEGV-----A 739  
QY 818 PKLALRETPALPVTSSSPSTTPNNMAWDFSSSTWSSSNEKETNNMDJMAAPSLTVPSA 877  
D 740 PVAAY-EAPVUAQVADT-----YNDNINT-----SSTPAA 768  
QY 878 GQLRQSAFTPATATGSSSPSVLQGEKEVGLQAQALYPMRAKKNHLEFKKSDVITLVE 937  
D 769 S-----ADULTAAG-----DV----- 778  
QY 938 QODMMWBEFVGQKGMFKSVYKLSGPRKSTSIDGPTSPASLKRVASPAKPAIPG 997  
D 779 ----- 778  
QY 998 EEFIAMTYEESQGLTLFOOGDVIIVTKKDGDMWTGTGVGKSGVPPSNVYRLKDSGSG 1057  
D ----- 778

D 779 EYIIAAYPEASEBGLDLSFSAEWMVVIKKEGEMWTGTIGRTGMPFSPNVYQAKDVGTA 838  
QY 1058 TAG-----KTGSLGKPELTAQVIAAATGPELTITAPQOLLIRKKNCGMWEGELO 1110  
D 839 TAAAPVESLQGMKAKSEIAQVIAPEATSTQSLTRQOLIMIRKKTDSGWEGELO 898  
QY 1111 ARGKKRQIGMFPANVYKLSLSPG--TSKITPTELPTAQAQAVC--QVIGMYDYTONDEL 1167  
D 899 AKGRRRQICMFPATVYKVLQGRNSGRNTPVSGSIEMTEQILDKVIALYRYKXONDEL 958  
QY 1168 AFSGQIINVLNKEDPDMMWKEFVSGVGLPSPNVY 1202  
D 959 SFDKDIISVLGRDEPEWWRGELNGLSLPSPNVY 993  
RESULT 2  
QYVIF7  
ID QYVIF7 PRELIMINARY; PRT; 1097 AA.  
AC QYVIF7;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
DE DAPI60 Protein.  
GN DAPI60 OR CG1099.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OK NCBI\_Taxid=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BERKELEY;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Celisnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Suton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champagne M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
RA Ballif J.F., Abhayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borrova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Crawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferlita S., Fleischmann W.,  
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodex A., Gong F., Gottrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam G.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mates J., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Mervilov G., Miliushina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacלב J.M.,  
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skipski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,  
RA Svaitkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of Drosophila melanogaster".  
RC Science 287:2185-2195(2000).  
CL -!- SIMILARITY: CONTAINS 4 SH3 DOMAINS.  
DR EMBL, AF003669; AAF53962.1; -.

QY	758	TIQGGDITAMVDESGTQTEPGLGELGKGTGWEPANAEKI	PENEVTPRAKPTVDTLSAPA	817
Db	694	TTFVGGDILIVPLFQNNPEPGLAGEINHGHWFPESV	VELLEVEGV-----A	739
QY	818	PKLARETPAPLPEVTSSEBPTTPNNADPSSSTWSSSNEK	PETDNWDTPWAOPSLTVPSA	877
QY	740	PVAAVE-----APVDAQVATVADTVN-----	DNINT-----SIIIPA	771
QY	878	GQLRQBSAFTPATATGSSSPFVLGQGEKVEGLQAOALY	WRAKKONHNLFNKSDVITYLE	937
Db	772	S-----ADLTAAAG-----	DV-----	781
QY	938	QQDMMWFGEVQGGKWFPKSYKLISGPVRKSTSIDTGP	TSASLKRVASPAKPAIPG	997
Db	782	-----	-----	781
QY	998	EEFIAMTYESSBOGDLTFQOGDVIIVTKKGDWMTGTG	VGKSGVFPSPNYVRLKD----	1055
Db	782	EYVIAAPYESAEEGDLSPFAGEMVWVVIKKEGMMVTG	ISRTGTFPSPNYVQKADVATAS	841
QY	1053	-----	-----SEGSG-----	105
Db	842	TAAAPVESLDOETTINGNAAYTAAPEVAEOEVOYOLP	PVQEPSEOPISSPGGAEEAHD	901
QY	1058	-----TAGT-----	-----GSLGKKEPIAQVIAVATGPEOLT	1087
Db	902	LDTVEGSIINTOSTKQSEPEASISYSPMSTSSMTPE	KRAKREIAOVIAIPYATSTBOLS	961
QY	1088	LAPQGLILIRKKNPGGWEEELQARCKRQIGFPA	NVYKLSPG--TSKITPTELPKTA	1145
Db	962	LTRQGLIMIKKTKDSGWBEELQAKRRRQIGFPA	TYKYVLQGGGNISGRNTPVSGSRLE	1022
QY	1146	VQPAVC-QVIGMTDYTHQNDDELAFSKGQIINV	LANKEDPMMKGEVSGVGLFPSPNYV	1202
Db	1022	MTEQILDKVIALPYPAQNDELSPDKDILSVLGR	DEPMMRGELINGLSGLFPSPNYV	1079

RESULT 3

061618

PRELIMINARY; PRT; 1094 AA.

AC 061618;

DT 01-AUG-1998 (T-EMBLrel. 07, Created)

DT 01-AUG-1998 (T-EMBLrel. 07, Last sequence update)

DT 01-MAR-2002 (T-EMBLrel. 20, Last annotation update)

DE Dynamlin associated protein isoform DAP160-1.

DN DAP160 OR CG1099.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxId=7227; [1]

RP SEQUENCE FROM N.A.

RC TISSUE=HEAD;

RA Roods J., Kelly R.B.;

RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.

CC -1- SIMILARITY: CONTAINS 4 SH3 DOMAINS.

DR EMBL; AF053957; AAC39138.1; --

DR HSPG; P29354; IGR1.

DR FLYBase; FBgn0023388; DAP160.

DR InterPro; IPR002048; EF-hand.

DR InterPro; IPR000261; EPS1-repeat.

DR Pfam; PF00036; ehand; 2.

DR Pfam; PF00018; SH3; 4.

DR PRINTS; PRO0452; SH3DOMAIN.

DR ProDom; PD000066; SH3; 4.

DR SMART; SM00054; EFh; 2.

DR SMART; SM00027; EH; 2.

DR SMART; SM00326; SH3; 4.

DR PROSITE; PS00018; EF\_HAND; UNKNOWN\_2.

DR PROSITE; PSS0002; SH3; 4.

KW SH3 domain.

SQ SEQUENCE 1094 AA; 120495 MW; 141370EB1FB960F0 CRC64;  
 Query Match 27.1%; Score 1698.5; DB 5; Length 1094;  
 Best Local Similarity 31.6%; Pred. No. 8,2e-72;  
 Matches 417; Conservative 201; Mismatches 329; Indels 371; Gaps 33;  
 QY 11 SLVVAITVEERAKHQOFLSKPIAGFTTGDQARNFFPFGSLPQVLAQIMALADNND 70  
 DB 4 AVDMAVAVPRERLKYOEQFALOPAGFVTGAQAKGFLOQLPPLILGQIMALADTSD 63  
 QY 71 GRMOVSESIAMKLIKXLOGVLPSTLPVWKQOPVAISGAPAFGIGIASMP---LT 127  
 DB 64 GKMMINETSILCKLNLKGMVFKVLPSSL-----LSL---TGDPVSMTPRSGSTS 113  
 QY 128 AVAPVPMGSIPIVGMSPPLVSVPPAAPPVPLANGAPVYIOPLPAPFAHPAAATWPKSSFSR 187  
 DB 114 SLSPLD---PLKGI-VPAVAVPVVPVAVPA--VATVISP-PGVSPSGGTPTSTN--- 162  
 QY 188 SGPSQNTKIQAKQSPFVSAAPRAEWAIVPQSSRLKXROLFNSHDKTMSGHLTGPQART 247  
 DB 163 --PSRHTSISERAPSTIESVN---QGEWAVOAQKRYTQVFNNDRTRSGYLTGSGARG 217  
 QY 248 ILMOSLPQAOLASIMNLSIDIDQDKLTAEFFILAMHLDIVAMSGOPLPVPVPEYTPPS 307  
 DB 218 VLVGSKLPQVTLAQIMTLSDIDGRLNCDEFILAMFLCEKAMAGKIPVTLTPOEWTPPN 277  
 QY 308 FRVAVSGSGMSVSISSSVDOFLPEPSSD-----EQPEKLP--VTEPKDK 353  
 DB 278 LRKTKSRG---SVSGVAVSRPGQAPASRHASVSQSGVGVADPADPAGJPGQTSFEDKR 333  
 QY 354 RENFERGVELLEKRRQALLEGQKOEERLAQERAEQEREREQOEQAEAROLELEKOLE 413  
 DB 334 KENVYKGAELDRRRKIMEDQOKREERERERERERERERERERERERERERERERER 393  
 QY 414 KORELERQEEERKEER 473  
 DB 394 ROREIEMEKER 453  
 QY 474 RKTLEFEALNDKKHOLEGLQDIPRCLATQROEISTNSRRLRLAETHLOOQOES 533  
 DB 454 NTQINVELSTINEKIKELSORICDTRAGVNTVKTVIDQMRTQRTDSMSESOQLARKEQ 513  
 QY 534 QOMTGRILPER-----QILSDQKOVQONSILHRDLSLTLKALEKEARQOLR 582  
 DB 514 NAKLQLOQBAAKWEAKSKASGALGGENAQEQ-----LNAAPHKQOLIINQIK 563  
 QY 583 EQLDEVERETSKLOEIDVFNNOULKELE-----IHSKQOLQKORSLEAPRLKQOEERK 637  
 DB 564 DKVENISKEISKEDINTNDVOMSELKALISALITKCEDLYKEYDVQRTSVLELKYNRK 623  
 QY 638 SLELEKQKEDQORVQERDKQWLEHVQOEQPRPKPHBEDRLKREDSVKKEAEERAKP 697  
 DB 624 NETSVSSAMDT-----GSSSM---EETGTTVTDY-----AVASNDLSALAAP 664  
 QY 698 EMODKQSKRLFPHQEPAKIATQAPWSTTEKGLPTISAQSVKVVYALVPFESRSHDEI 757  
 DB 665 AVD-----LGGPAP-----EG-----FVKYQAVEFNARNAEII 693  
 QY 758 TIQGGDIYVWESQTEGFGWLGSELKGTGTFPFPANYAEKIPENVPVPAPRVDTLTAPA 817  
 DB 694 TFEVGGDIILVLEQNAEPEGWLAGINGTGFPEPVYEKLEVGEV-----A 739  
 QY 818 PKLARETPALPYTSSPSTTPPNWADFSSSTWSSSNEKETNMOTWMAQSPSLTVPASA 877  
 DB 740 PVAIV-EPAPVAVQADT-----YNNINIT-----SSTPAA 768  
 QY 878 GOLRORSAPFTATGSSPSVPLQGEKEVGLQAQALYPMWRAKKNHLNFKSDVITVLE 937  
 DB 769 S-----ADLTAAG-----DV----- 778  
 QY 938 QQDMMWVEGEVGGQGWPFKSYVKLISGPVRKSTSIDTPTESPASLKRVASPAKPAIPG 997  
 DB 779 ----- 778

QY 998 EEFIAMTYESESQDGLTFQOQDVIVTKKDDMWGTGVGDKSGVFPSPNVRLKND----- 1052  
 DB 779 EYVIAVPPESAEEDLFFSAGEMVMVKKGEWMWTGTIGRTGMFPSPNVYKXADVGTAS 838  
 QY 1053 -----SESSG----- 1057  
 DB 839 TAAEPVESLDQETTLNGNAAYTAAPVEAOBOVYQPLFVQEPSSOPISSPVGAEEAHED 898  
 QY 1058 -----TAKT-----GSLKKEPEIAQVIAVYAAQTEPOLT 1087  
 DB 899 LDEVQINTQKTSSEPPAESYSPMRSTSSMTGMAKSELSIAQVIAPEATSTBOLS 958  
 QY 1088 LAPGQILIRKKNPGCWMEGELQARQKRGQIGWPPANYKLLSPG--TSKITPTELPKTA 1145  
 DB 959 LTRGQILIMRKKTDSGWMGEGELQAKGRROIGWFPATYKVLQGGRNSGRNTPVSGSRTE 1018  
 QY 1146 VQPAVC-QVIGWYDYTAQNDDELAFSKQIINVLNKEDPDMWKGEVSGQVGLFPSNVY 1202  
 DB 1019 MTEQILUDKVIALLYPYKAQNDDELSPDKDITISVLAGRDEPEWWRGELNLSGLFPSNVY 1076  
 RESULT 4  
 ID 09U2T9 PRELIMINARY; PRT: 1097 AA.  
 AC 09U2T9,  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
 DE Y16A8C.36 protein.  
 GN Y16A8C.36.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Pelodetinae; Caenorhabditis.  
 OX NCBI\_Taxid=6239;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA McMurtry A.A.;  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C. elegans: A platform for  
 RT investigating biology";  
 RT Science 282:2012-2018(1998).  
 CC -1- SIMILARITY: CONTAINS 5 SH3 DOMAINS.  
 DR EMBL; AL117204; CAB5138.1; -.  
 DR HSSP; P29355; ISEM.  
 DR InterPro; IPR000194; ATPase a/bcentre.  
 DR InterPro; IPR002048; EF-hand.  
 DR InterPro; IPR000261; EPS15\_repeat.  
 DR InterPro; IPR000108; Neu\_cyt\_fac2\_2.  
 DR InterPro; IPR001452; SH3.  
 DR Pfam; PF00036; ehand; 2.  
 DR Pfam; PF00018; SH3; 5.  
 DR PRINTS; PR00459; P67PHOX.  
 DR PRINTS; PR00452; SH3DOMAIN.  
 DR ProDom; PD000066; SH3; 5.  
 DR SMART; SM00027; EH; 2.  
 DR SMART; SM00326; SH3; 5.  
 DR PROSITE; PS00152; ATPASE\_ALPHA\_BETA; UNKNOWN\_1.  
 DR PROSITE; PS00018; EF\_HAND; UNKNOWN\_2.  
 DR PROSITE; PS00002; SH3; 5.  
 KW SH3 domain.  
 SQ SEQUENCE 1097 AA; 122072 MW; 1C2BA5F103968372 CRC64;  
 Query Match 21.3%; Score 1333; DB 5; Length 1097;  
 Best Local Similarity 29.4%; Pred. No. 1.1e-54;  
 Matches 369; Conservative 200; Mismatches 454; Indels 234; Gaps 39;  
 QY 15 VAITVEERAKHQOFLSKPIAGFTTGDQARNFFPFGSLPQVLAQIMALADNND 67  
 DB 15 VAITVEERAKHQOFLSKPIAGFTTGDQARNFFPFGSLPQVLAQIMALADNND 67

```

Db 5 WEVSDAEYCKNFMF-----GQLTGGQPFMAVTAARNALMRSNLPVOVLSQIMALSUL 57
Qy 68 NNDGRMDQVEFSIAMKLIKLIKQYQSTLLPVYMQQVPAISSAPAFGIGIASMPULL 127
Db 58 DKDGRDLIREYSIAMRLAINCLAGIPPIPOLPSSLVPA--RNAPPTMGs----- 107
Qy 128 AVAPVPMGSLPVVGMSPVL-----SSVPAVPLPANCAPVITQPLAFAPHAATWIK 181
Db 108 -----RHGSVDYSQTLPEPAIDRRMSQSYIPSA--PVSIASTP----- 142
Qy 182 SSSFSRSFGSSQNTKLQKQSFVVASAPPAEMAVPQSSRLKYROLFNSHDKTMSGHLT 241
Db 143 SSSHNSTISAGSPLN-----NDRNVEFGROLBNWALPHNKKLYSOLFALNDEKRGIS 196
Qy 242 GPQARTILMOSSLPQOALASIMWLSIDODGKLTAEFFILMHLIDVANSQGPLPVLP 301
Db 197 SQVGRSALGIGSLPTVNLAMIFLSDVYKDGKLSVDEYALISQWIMFPGSGLPRTITL 256
Qy 302 EYIPPSFRVRSSGMSVSSSVQDRLPREPSSEDEQPEKLPVTPEDKKRENERGS 361
Db 257 ELY-----RMCGISSRSANNTPELEPGAEPPQ--KSPAPYTFEDKRODNLKSGQ 303
Qy 362 VELEKRRQALIEOORKEOERLAERAEORKEBEROERQAEKOLELEKOLEKOLEBERO 421
Db 304 ABLERRRRVLEEBEORRAVEKKEERBEAKKURBEROERQAEVEROELERQRLIEBO 363
Qy 422 REBE-----RKREIR-----EAKRELE-----ROROLEMERNRROELLNORKEQETV 469
Db 364 REEBEKKRRLEEMERRREDEKMRKVOMKAKVOMQVOMQEQKMMFVYQKQENRMLQ 423
Qy 470 LKARRKLELEBALNDKKHOLEGKQODIRCRATOROEIESTNKSRELIAETITLQOO 529
Db 424 ROOREKTLQFOLADEKVIDEVDIGKAKEVAEVTGFIERRSTRDEKVARI-----XE 479
Qy 530 LOSQOMLGRLEIKOLISDOLKOVQONSILHRSLSLTLKALEAKLAROUEOLD----- 586
Db 480 LQSTNQ-----KTAIESQELGHQ--LQKQSAHKEI--TQKSELALRRKDAIKALEEDA 533
Qy 587 EYERETRSLQOEIDVFN--NNQLKELREIHSKQOLKORSLAARLKQKEOERKSLERQ 644
Db 534 LELSTEREKSYNTQELIKTKKEKYKTQVSK-----LVAKREBYRNSFEL----- 578
Qy 645 KEDAOORRVERDKOMLEHVOOEOPREPRK-----HEEDRLKREDSVK 688
Db 579 -----LVHAQTHASKIGEFPAKSAASAAPAPAPARTTTNGFANFNDAFGEFD--K 631
Qy 689 KEAEERAKPEMOKOSLFLPHROEPKALQIOWSTTEKREPLTISAOESKVVYVYALP 748
Db 632 TDASQRFDAFGATST-----ADPFAQIAQAP--AHSKGAVDQSAFNIDHYKCRALPA 683
Qy 749 FESRSHDEITQGDIVWDESQTEGPEGMIGELKGTGMFPANVAKI PENEVPTPAP 808
Db 684 FEARSEBELSFEGDVIIVFOSHAEFGMRAGOLREKVGFPFAFVBAI--AVPTPG-- 739
Qy 809 VTDLTSPAKKALRETPAPLPVTSSEPTTPNNMADFSSSTWSSSNEKETDNMTWMA 868
Db 740 -----GDPRIQNMPEMNTPESSVDQIV--KAARKEI-----AA 772
Qy 869 QPSLTVPSAQLOKRSAPFTPATATGSSPVLQGEKVEGLQOALYPMWAKKONHLNEN 928
Db 773 AMGLTEGCA-----PPASSAPAAVAIS-----QCIAGFOWARNEEDLSIFA 814
Qy 929 KSDVITVLEQOQDMWFG--EYOGQKGMFPKSYVVLISGPVRSKTSITGTPESPASLSKVA 987
Db 815 KQDTIEVLEQEKMKGRNPAIGEIMFPKSYVKEVAGTSTTPTI--VSPKASAGAPGAA 873
Qy 988 SPAKPAIPGE-----FIAMYTYSSEQDILFQOGDVIIVTKKQDMWMTG 1034
Db 874 AGAQYDVVPDVTLOASETAPQOOLYVIVDFEAVETTDLALHVGDITLLEKNDMMWG 933
Qy 1035 TVGDKSGVFPSSNYRLKDSGSGTAGKT--GSLGKPEI--AQVLAASYATPEQULTLA 1089
Db 934 RCGREGIFPANYVEISVQAGDPTPTQAPTPAAPPTVLCBAKVVDVFAAPNOLGIK 993

```

```

Qy 1090 PQGLILIRKKNPGWBGSELQARGKROIGWFPANVYKLSPTGSKITPTPELEKTAVOPA 1149
Db 994 VGEIYKIREKSAAGWBEGLIRNGKP--IAGWFPGEYKVLVEEASAPAT----- 1040
Qy 1150 VQOIVGMQDYTAQNDDELAFSGQIINVNLNKEDPDMWG--EVSQGVGFPSNYVK 1203
Db 1041 --RATAVDYEASQPDDELGFRTGDVITVTKSEAMWMSGHREDDPSKSGLFPSSNYQ 1095

RESULT 5
08T068
ID 08T068 PRELIMINARY; PRT; 662 AA.
AC 08T068;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE LD23686P.
GN DAPI60.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Empidoidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Fartan D., Friese E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuno J., Paclio J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.B., Rubin G.M., Celinker S.,
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY069517; AAL39662.1;
SQ SEQUENCE 662 AA; 72553 MW; 3532CB7034BD7AC CRC64;

Query Match 12.7%; Score 794.5; DB 5; Length 662;
Best Local Similarity 25.9%; Pred. No. 1e-29;
Matches 221; Conservative 108; Mismatches 203; Indels 321; Gaps 21;

Qy 459 QNRKQEGTVLKKARKTLEFELALNDKKHOLEGKQODIRCRATOROEIESTNKSREL 518
Db 4 QEROERVLKQKANTQJNLVELSTLNEKIELSQRICDTRAGTAVTKVITDGMRTQRTD 63
Qy 519 RAETTHILOOQLOESQMLGRLEPEK-----QILSDOLKOVQONSILHRSLSLTL 567
Db 64 SMSMSQLKAIKEQNALLOLTOERAKMEKSAAGALGEMAOBO-----L 113
Qy 568 KRALFAKELAQOLREOLDEVERETRSLQOEIDVFNNOQKELRE-----IHSKQOLQOR 622
Db 114 NAAFAHQOLIINOIKDKVENISKEIESKEDINTNDVOMSELKALSLITKCEDLYKEY 173
Qy 623 SLEAKRLKQKEOKSLELEKQKEDAOORRVERDKOMLEHVOOEOPREPRRPHNEEDLKR 682
Db 174 DVQRTSVLELKYNRKNETSVSAMD-----GSSASW-----EETGTVTVDPY----- 216
Qy 683 BDSYKKAEBERAKEMQDQOSRLPHROEPKALQIOWSTTEKREPLTISAOESKVVY 742
Db 217 --AVASNDISALABAVD-----LGGPAP-----EG-----FVK 243
Qy 743 YRALYFESRSHDEITITQGDIVWDESQTEGPEGMIGELKGTGMFPANVAKI PENEV 802
Db 244 YQAVEYFARNABEITFVPGDITLVPLEQNAPEGMILAGEITNGHTGWFPESEVLELEVGEV 303
Qy 803 PTPAPVTDLTSAKPAKALRETPAPLPVTSSEPTTPNNMADFSSSTWSSSNEKETDN 862
Db 304 -----APVAABE--APVDAQVATVADTYN-----DN 327
Qy 863 WDTMAQPSLTVPSAQLOKRSAPFTPATATGSSPVLQGEKVEGLQOALYPMWAKKD 922
Db 328 INT-----SSTPAPS-----ADLTAG----- 344
Qy 923 NHLNENKSDVITVLEQOQDMWFGVEVQOGKGMFPKSYVVLISGPVRSKTSIDTPTESPAS 982

```

```

Db 345 -----DV----- 346
Qy 983 LKRVASPAKKAIPGEFIAMTYESSHQGLTFQCGGVIVVTCKDDMMWTGTVGDSGV 1042
Db 347 -----EYIIAIPYIESAEGLSFSAGEMVWVIKKEGEMTGTIGSTGM 391
Qy 1043 FPSNVYRLKD----- 1052
Db 392 FPSNVYQADVGTASTAAEPVESIDQETTLNGNAAYTAPAEQGVYQPLPVQEPSEQ 451
Qy 1053 ---SEGS-----TAGT-----GSLKKEPIA 1072
Db 452 PISPGVGAEEAHEDLDTEVSQINTQSGSEPAEYISRPMSSTSMTPGMRKASEIA 511
Qy 1073 QVIASVATGEQULIARQGLILIRKKNPGWEGELQARGKRQIGFPPANYKLLSPG 1132
Db 512 QVIAPYEATSTEQSLTTRGQILIMIRKKTDSGMEGELQAKGRROIGMFPATYVKVLQGG 571
Qy 1133 --TSKITTELPRKTAQVAVC-QVIGMYDYTAQNDDELAESKQIINVLNKEDPMWKE 1189
Db 572 RNSGNTFVSGSRIEMTEQIILDKVIALYPKAQNDELSPDKDIIISVLGRDEBEMWGE 631
Qy 1190 VSGQVGLFPSNVY 1202
Db 632 LINGSLGFLFPSNVY 644

RESULT 6
Qy 09UBC2 PRELIMINARY; PRT; 864 AA.
ID 09UBC2
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Epidermal growth factor receptor substrate Eps15r.
GN Eps15r.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RA Naeashima S., Morinaka K., Ikeda M., Kishida S., Koyama S.,
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RA Ueki N.;
RL "HRI NTT human fetal brain cDNA project.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF110265; AAF21930.1; -.
DR EMBL; AB015346; BAA88118.1; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000261; Eps15_repeat.
DR Pfam; PF00036; ehand; 4.
DR SMART; SM00054; Efh_3.
DR SMART; SM00027; Eh_3.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
KW Receptor.
SQ SEQUENCE 864 AA; 94254 MW; FA126069F6E00387 CRC64;

Query Match 8.6%; Score 536.5; DB 4; Length 864;
Best Local Similarity 23.5%; Pred. No. 1.7e-17;
Matches 216; Conservative 139; Mismatches 349; Indels 215; Gaps 28;
Qy 15 WAITVEERAKHDOFLSLKPIAGFITGDQARNFFQSGLPQVLAQIUALADNMNDGRMD 74
Db 121 WAWRVEERAKHDOFLSLKPIAGFITGDQARNFFQSGLPQVLAQIUALADNMNDGRMD 180
Qy 75 OVERSIANKLIKLOQIGTQPLSTLPVVKQOPVAISSAPAGIGSIASMPPLTAVAPVM 134

```

```

Db 181 REFVAMHLVYRALEKEPVSALEPPL-----IPSKRKTVFP 220
Qy 135 GSIPVNGSPPLVSVPPAAVPLNAGAPVYIOLPAPAHAAATWPKSS--SFSRSGGS 192
Db 221 GAVPLPASPFP-----PKQSLASTPSHSGVS 246
Qy 193 QUNTUKLQKASFDVASAPPAEMAVPOSSRLKYROLFNSHDKTMSGHLTGRQARTLMQS 252
Db 247 SLNLSGSSLPKSLKQOTQPTVMVWVVPVADKMRFDEIFUKTLIDLIDGVYSGQEVKEIFMHS 306
Qy 253 SLPOAGLASIWNLSIDDDGKLTAEPIFLAMHLIVAAVS--GQPLPVYPPYIPSPFRV 311
Db 307 GLTONLHAHTMALADTRQTKLSKQFALAMFYIOAKSKGIDPPQVLSPMVPPSESGT 366
Qy 312 RSGSGMSVSSSSVDQRLPEEPSSDEQOPEKKLPVTFEDKKRENFERSVLEKRRQAL 371
Db 367 PGPDSSGSLSGSEF-----TGVELELDISQ-- 391
Qy 372 LEOQRKEQERLQLERAQERKERERQQAQKQULELEKOLEKOLEKOREBERRKEIE 431
Db 392 -----ELAQOREKYSLEQDIREKEEAIROKTSVQ--ELONDLRETSLOELEAQ 441
Qy 432 RREAAKR--ELERQQLERWRNRQELNORNKEQEGTVV--LKARKKTLFEFLAALND 486
Db 442 KODAQDRIDEMDQOK-----AKLRMLSDVQKQODETOMISLKTQIQOSQESDLKSGED 496
Qy 487 KKHQLEGLKQIDIRCLARQOEIESTNKSRELRLAETHLQOQLOESQOMGRILPEKQI 546
Db 497 DLNRAKSELNRLQOEBETOLESIOGRVQLETIILSLKSTQDEINQASKLSQLESKOE 556
Qy 547 LSDQLKQVQVQ--NSLHRDSILTLKRALFAKELARQQLREQLDEVRETRSKLOEIDVFN 604
Db 557 AHRSLERQYQVLDGAHAGSLTDLANLSEGVSLAERKSGAND-----DPRKN 603
Qy 605 QULKELEIHSKQOLOKORLEAARLKQKOEERKSLELEKOKEDAGRVOERDKQMLEHVQ 664
Db 604 --KALLFSNNQELHBDPF-----QTEBDPKS--DPRKGAIDPRFGDQFQNDP-----F 647
Qy 665 QEBQPRKPHEDBKLEDSYRKKEAEERAKREMQD-----KOSRLFNHQERAK 715
Db 648 ABOQTSTTDPFGGDPRKSDPRGSATDPFKKQYKNDPFTSDPRTKPSL-----PSK 701
Qy 716 LATQAPWSTTEKGPLTISAOESVKVYVYALYFPESRSHDEITITOPGDIWVDESQTER 775
Db 702 L---DPFESSD--PFSSSSVSSKSGDPFTLDPFGSGSFSNAB-----GFADPQWMSKP 750
Qy 776 -----GWLGEELKAGTGMFPANTAEKI PENEVPTPAKPVYTLTSAAPKIALRETPAP 828
Db 751 PPSGPFSTSLGG-----AGPSDDPFKSKQDTPALP-----PKKAP-----PR 789
Qy 829 LPTVTSSEPTTNMNAQFSSSTWSPSSNEKPE-----TDNMDTMAQOSLTVPSAGQLRQ 882
Db 790 KPPSGKSTVSQLGSHDPF-----BAQDFQGLAGDSGDPFGSKGFGDPRFG----- 837
Qy 883 RSAFTPATATGSSPSPVLG 901
Db 838 KDPFVPSAAXPKSASASG 856

RESULT 7
Qy 08WQ61 PRELIMINARY; PRT; 1253 AA.
AC 08WQ61;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Eps-15 protein.
GN Eps-15.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.

```



[illegible]

Db 708 QRPFGSDA-----FRANKSNAIITPEPKDDFGSGPPAALHAFTGQ----- 748

Qy 780 GELKGTGMPFANVAEKIPENEVPPAKVPTDLTSAAPKALRETPAPLPVTSSEP---- 836

Db 749 GQVLSFN-----AQSGGPPRPESPSPALP-----PKSKVPPRPAPPPAAOPTG 795

Qy 837 --STTNMMADF-----SSTWSSSNEKEPETDNDMAAQSLVLP----- 875

Db 796 FGSGGGGAGPDEPDFDNNKLHIIPSAPSAVALSP-----LPPTLPAPLPVVS 846

Qy 876 -----SAGLKRSAFTP-----ATATGSSPSPVLP-----GQGE 904

Db 847 LLDSTFLPDDPQGIHQAASTPNPPIVPTVHTLLQTSLSTPAAPBALASTSGSG- 905

Qy 905 KVEGLQA-----QALYPMRAKDNHLNFKSDVITVLEQDMMPF 944

Db 906 SVAGAGAHLPSSVTITTPATSLNQHLSNSNPPLQNTQSTADVKLETKAVVSVFDA-----F 960

Qy 945 GEVQGGKGMFPKSYVVLISGPFRKSTSIDTGP-----TESASLSKRAVSAAPALPGE 999

Db 961 GEIGTRKAPTP-----SLTGP-----TDFKDDPFKDYREYDEPFSIK--DFAB--EGEE 1006

RESULT 8

Q60902 PRELIMINARY; PRT; 907 AA.

AC Q60902;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE Eps15r protein.

GN Eps15-RS OR Eps15R.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxId=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BALB;

RX MEDLINE=96003812; PubMed=7568168;

RA Wong W T., Schumacher C., Salscini A.E., Romano A., Castagnino P., Pelicci P.G., Di Fiore P., "A protein-binding domain, EH, identified in the receptor tyrosine kinase substrate Eps15 and conserved in evolution.", Proc. Natl. Acad. Sci. U.S.A. 92:9530-9534 (1995).

RL EMBL: U29156; AAA87202.1; -

DR MGD; MGI:104582; Eps15-RS.

DR InterPro; IPR002048; EF-hand.

DR InterPro; IPR000261; Eps15\_repeat.

DR Pfam; PF00036; efhand; 4.

DR SMART; SM00054; Eph; 3.

DR SMART; SM00027; EH; 3.

DR PROSITE; PS00018; EF\_HAND; UNKNOWN.1.

DR SEQUENCE 907 AA; 99320 MM; 2FA79DFB4C834E1E CRC64;

Qy Query Match 8.4%; Score 527; DB 11; Length 907; Best Local Similarity 22.7%; Pred. No. 5e-17; Matches 225; Conservative 134; Mismatches 362; Indels 270; Gaps 30.

Qy 15 WAITVERAHHDOQFLSLPIAGFITGDOARNFFQSGLPQVLAQIWMALDNNNDGRMD 74

Db 121 WAVEVEEKAQFDGIFESLLPVNGLSDGKVPKVLNMSKLPUDVIGRWVMDSDIDKGDHLD 180

Qy 75 QVESIMAKLIKLTQGLQYQLPSTLPPVKKQDPVALISSAPRGIGISMPPLTAVAPVEM 134

Db 181 RDEFAVVMHLYVRALEKEPVPSTLPPPL-----IPSSRRKKTVEA 220

Qy 135 GSIVGVMSPEPLVSVPPAAVPLPANGAPVVIQPLPAFAHPAATWPKSS--SFSRSGPGS 192

Db 221 GAVVPLPASP-----PQDSLRSTPSHGVS 246

Qy 193 QLNKLQKAGSPFVASAPPAAEMAVQSSRLKTYROLFNSHDKTNSGHLTGPQARTILMOS 252

```

Db 247 SLNSTRGSLSPKHSV-KQPPVA-WVVPVADKKRPDEIFLKTDLDDGYSGGEVKEIFMHS 304
QY 253 SLTPQALASIMNLSIDDDGKLTAEFTLAMLIDVAMS-GQPLPVLPEVPIPSFRV 311
Db 305 GLTQNLALHIALADTGTGKLSKEQFALAMYFIQKQSKGIDPPQVLSVPMVPSRG 364
QY 312 RSGSGMVISSSSVQRLPEPPSSSEDOQPEKLPVTFEDKKRENFERGSELEKRRQAL 371
Db 365 PIPSSSTLAGEF-----TGVEELDDISG-- 389
QY 372 LEQAKREORLAQLEPRAQEKREKREKREKQLELEKQLEKQRELEKREKREKRE 431
Db 390 -----EIAQLOREKYSLEODIREKEEAIKQKTSFVQ-----ELQNDLD 427
QY 432 RREAKRELEKROLEWERNRROELLNQRNKEQEGTVLKKARKTLEFELEALNDKHOL 491
Db 428 RETSSLOLEKQKQDAQR-----LDMDQQA 457
QY 492 ESKLQDIRCLATQROEISTNKSRELIATHLQOQLOESQOMLGRLPEKQILSDQ 551
Db 458 RDLMSDVOKOCODETQITSLTKTOIOSQESDLKSGEDDLNRAKSELNRLQEE----TOL 513
QY 552 KQVQONSLSRSLTLTKALEAKELARQLEQDDEVRETRSKLQETIDVFNNGKLE 611
Db 514 EQ-----SIQAGRAQLETTILRSUKTQDDINQARSKLSQLE 550
QY 612 IHSKOQLKORSL-----AARLKQEKREKSLLEKQEDAQRRVQEKQOML- 660
Db 551 SH-----LEAHSLEQYDVPDGVSTSLPLDLATINEGILLERGFAMDPFNKXALLF 606
QY 661 EHVQOEOPRP--RKPIEDRLKREDSVRKKEAEERAKPEMOP-----KQSR 706
Db 607 SNNSELELPDPOAEKDPKSDPFKADPFKDPFQSDPFSSQQAATDPFGSDPFKSDP 666
QY 707 FHPHQPAPKATQAPWSTTEKQPLTISAQESVYVYVYALPFPESRSHDETTIQGIV- 765
Db 667 FHSSSDPFKQKQTDNFTSDPFTKNSLPSKLDPFSSSPSSSSISKSGSPFGTLD 726
QY 766 -----WDESGTGED--GWLGEIKGKTGMFANVAEK-----IPENEVPTP 805
Db 727 PFGSSSFSAEGFADFSQMSKRPPEGPFSSSL-GOTGFSDDPFKSKQDTPALPKK-PAP 784
QY 806 AK-----PVDTLSAPAPKALARETP--APLPVTSSEPTTPNMAD-SSSTMP- 851
Db 785 PRPKPSSQSTPVSQSGSDPF-----ESPDPQPLGADSGEPQNKKGfGedpSGADPF 839
QY 852 -SSSNEKETDNDTWAQPSLTVPSAGOLRQSAFTPATATGSPFVLQGEKEVGLQ 910
Db 840 APSSSAK-----PKASS-----SGFADFTSGFNEQQLAMAKSEKAE 879
QY 911 AQALYPWAKKDNHLNFKSDVITVLEQDM 941
Db 880 QERLARLRROEDL-----ELATALSKADM 905

RESULT 9
Q23418 PRELIMINARY; PRT; 751 AA.
AC Q23418;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Hypothetical 81.8 kDa protein.
GN ZK1248.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peleoderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;

```

```

RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT Investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Latreille P.;
RL "The sequence of C. elegans cosmid ZK1248.";
RN [3]
RP Submitted (JUN-1995) to the EMBL/Genbank/DBJ databases.
QY 372 LEQAKREORLAQLEPRAQEKREKREKREKQLELEKQLEKQRELEKREKREKRE 431
Db 390 -----EIAQLOREKYSLEODIREKEEAIKQKTSFVQ-----ELQNDLD 427
QY 432 RREAKRELEKROLEWERNRROELLNQRNKEQEGTVLKKARKTLEFELEALNDKHOL 491
Db 428 RETSSLOLEKQKQDAQR-----LDMDQQA 457
QY 492 ESKLQDIRCLATQROEISTNKSRELIATHLQOQLOESQOMLGRLPEKQILSDQ 551
Db 458 RDLMSDVOKOCODETQITSLTKTOIOSQESDLKSGEDDLNRAKSELNRLQEE----TOL 513
QY 552 KQVQONSLSRSLTLTKALEAKELARQLEQDDEVRETRSKLQETIDVFNNGKLE 611
Db 514 EQ-----SIQAGRAQLETTILRSUKTQDDINQARSKLSQLE 550
QY 612 IHSKOQLKORSL-----AARLKQEKREKSLLEKQEDAQRRVQEKQOML- 660
Db 551 SH-----LEAHSLEQYDVPDGVSTSLPLDLATINEGILLERGFAMDPFNKXALLF 606
QY 661 EHVQOEOPRP--RKPIEDRLKREDSVRKKEAEERAKPEMOP-----KQSR 706
Db 607 SNNSELELPDPOAEKDPKSDPFKADPFKDPFQSDPFSSQQAATDPFGSDPFKSDP 666
QY 707 FHPHQPAPKATQAPWSTTEKQPLTISAQESVYVYVYALPFPESRSHDETTIQGIV- 765
Db 667 FHSSSDPFKQKQTDNFTSDPFTKNSLPSKLDPFSSSPSSSSISKSGSPFGTLD 726
QY 766 -----WDESGTGED--GWLGEIKGKTGMFANVAEK-----IPENEVPTP 805
Db 727 PFGSSSFSAEGFADFSQMSKRPPEGPFSSSL-GOTGFSDDPFKSKQDTPALPKK-PAP 784
QY 806 AK-----PVDTLSAPAPKALARETP--APLPVTSSEPTTPNMAD-SSSTMP- 851
Db 785 PRPKPSSQSTPVSQSGSDPF-----ESPDPQPLGADSGEPQNKKGfGedpSGADPF 839
QY 852 -SSSNEKETDNDTWAQPSLTVPSAGOLRQSAFTPATATGSPFVLQGEKEVGLQ 910
Db 840 APSSSAK-----PKASS-----SGFADFTSGFNEQQLAMAKSEKAE 879
QY 911 AQALYPWAKKDNHLNFKSDVITVLEQDM 941
Db 880 QERLARLRROEDL-----ELATALSKADM 905

Query Match 8.1%; Score 506.5; DB 5; Length 751;
Best Local Similarity 24.7%; Pred. No. 3.7e-16;
Matches 214; Conservative 115; Mismatches 273; Indels 263; Gaps 34;
QY 15 WAITVERAKHQQFSLKPIAGFITGQARNFPGQGLPQVLAQIWMALAMNNNGMD 74
Db 111 WPTTPDQAKYVSTFQSLNPNVNGKLSGAHVRYVLMNSGLDAHALRIWELSDQKDGSLD 170
QY 75 QVEFSIAMKLILKLYGYQLPSTLP--VMQOPVAISSAPAFGIGIASMPRLTAVAV 132
Db 171 RIEMSVALLHYVRSIQSPVPAPQLPRLNLIHSEKAMVAHSSSNF-----AAP--HPP 221
QY 133 PM-----GSIPVGMSPPLVSVPPAAYPLANGAPVYIOLPAPAHNA----- 176
Db 222 PMWGASGAVTSLDDNNSSQ--YSATMPRAYP-----QPRATISAQAHVNGST 271
QY 177 -ATWPKSSFSRSGQSOQLNTKLQAGSFDAVAPAEAWVPOSSRLUKYQLFNSHDT 235
Db 272 GASTPISASHS-----IHSF-----PAGEWPIINTGD--YADQFAQTDTN 308
QY 236 MSGHLTGPAKRTILMOSSLPQALASIMNLSIDDDGKLTAEFTLAMLIDVAMSGPL 295
Db 309 KQGLVDGMRAPMWTTGISAQILAHWALADIKKCGQLNLEQFALTMHLLDMARKGESI 368
QY 296 PVLPEVYIPFSFRVRSQSGSVSSSVDOQLPEEPPSSSEDOQPEKLPVTFEDKKRE 355
Db 369 PESTPLHLIPFSFRPTESALHHPAQSVSTPOLPEATSMELKELF-----GE 417
QY 356 NFERGSELEKRRQALLQERKEORLAQLEPRAQEKREKREKREKQLELEKQLEKQ 415
Db 418 NEF--WKQLAESIQMIVERTKAEAVQLF--ADMTINSSIK--NIQVELATLSTV 470
QY 416 RELEROREBERK-----EIERREAKRELEKROLEWERNRROELLNQRNKEQEGTV 469
Db 471 KQERQKGEATRLADYDTQIOLESA--CAQKQETEDTEKRNQOJIDEDAKNAD--- 524
QY 470 LKARRTLFELEALNDK--HLEGLQDIRCLATQROEISTNKSRELIATHTHQL 527
Db 525 CKA-----NDEKEMELKKEIEMLDNQKTYRGELVKEVTSQREQVAVLT--- 569
QY 528 QOQLOESQOMLGRLPEKQILSDQLKOVQONSLSRSLTLTKALEAKELARQLE-EDL 586
Db 570 -----TLERKE-ARDQIOMEXLD 586
QY 587 -EYERETRSKLOEIDVFNNGKLEKREIHSKQOLQKORSLEAARLKQEKREKSLLEKQ 645
Db 587 AATENTTKLTBOVSADAVKSEBEKQI-----LRSQQLSLIVIQSLSDDTVYGATG 641

```

```

QY 646 EDQRRVQERDKQWLEHYQOEOPRRPKRPHEDRLKREDSVYRKKAEERAKPEMODKQR 705
DB 642 TSSQ-----NHVQOPDP-----FASARAPAD----- 665
QY 706 LFRPHQEPKATLQAPMSTTEKGLTISAQESVYVYRALYPFESRSHDEITIQPDIV 765
DB 666 -----FPAQVDQ-----FGSSGH----- 678
QY 766 MDESQTEGPGWLGELKGTGTFPAN--YAEKIPENEVPTPAKPVTDLTSAAPKTLAR 823
DB 679 -FDAAFPTDPFAQGG-----FPSDSGFAQSAAPAK-----PAP-----RPAPKSAR 719
QY 824 ETPAPLPTVTSSE-PSTTPNMADFS 847
DB 720 ETPVNDPFAPBSQGSTOPAGFADFA 744

RESULT 10
Q9BIF4 PRELIMINARY; PRT; 796 AA.
ID Q9BIF4;
AC Q9BIF4;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE EHS-1.
GN EHS-1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=21376507; PubMed=11483962;
RA Salani A.E., Hilliard M.A., Croce A., Arbucci S., Luzzi P.,
RA Tacchetti C., Danielli L., De Camilli P., Pelicci P.G., Di Fiore P.P.,
RA Bazzicalupo P.,
RA "The Eps15 C. elegans homologue EHS-1 is implicated in synaptic
RT vesicle recycling."
RL Nat. Cell Biol. 3:755-760(2001).
DR EMBL; AY027560; AAK13051.1; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000261; Eps15_Repeat.
DR Pfam; PF00036; ehnd; 3.
DR SMART; SM00054; EFh; 2.
DR PROSITE; PS00018; EF_HAND; 5FAE9474EC258E98 CRC64;
SQ SEQUENCE 796 AA; 86644 MW; 5FAE9474EC258E98 CRC64;

Query Match 8.1%; Score 506.5; DB 5; Length 796;
Best Local Similarity 24.7%; Pred. No. 4e-16;
Matches 214; Conservative 115; Mismatches 273; Indels 263; Gaps 34;

QY 15 WAITVERAKHDQFLSKPIAGFITGDQARNFFQSGLPQPVLAQIWMALADNMNDGRMD 74
DB 156 WPTPTDQAKYDSIFOSLVNNGKLSGAHYRVLNMSGDLDAHALARIWELSDQDKXGND 215
QY 75 QVEFSIAMKLKIKLQGYQLSTLP--VMKQOPVAISSAPAGIGISMPPLTAVAV 132
DB 216 RIMSVYALHLYVNSLOSQDPVPAQLPRLIHPSKAMYAHSPN-----AAP--HPR 266
QY 133 PM-----GSIPVVGMPPLVSVPPAAPPVPLGAPPVIOPLFAFHPA----- 176
DB 267 PMMGSRAGSVTSIDVNMGSQ--YSATMPRAYP-----QGRAYSAQAHVNSRPS 316
QY 177 -ATWPPSSSRSGPSGSQLTKLQKQSPVVASAPPAEVAVQSSRLKYRQLFNSHDT 235
DB 317 GASTPISASHS-----IHSF-----PAGEWPINTGD---YAOQFQDTN 353
QY 236 MSGHLTGPQARTILMOSLPOAQLASIMNLSDIDQGLTAEFIAMHLIDVAMGQPL 295
DB 354 KQGLVGDGMRAKRAFMPTTGLSAQILAHYMALADIKKCGQLNLEQFALTMHLIDMAKGESEI 413
QY 296 PVLAPPEYIPPSFRYRVSQSGMSVSSSVQRLPEEPSEDEQPEKPLPYTFEDKRE 355

```

```

DB 414 PSELPLHILIPPSFRPPPEPSALHHPQSVSTPQLPEATSMIEKALE-----GE 462
QY 356 NPERGSVELERKQALLQOEKREORLQERAEQERERERQOEAKROLLEKOLEKQ 415
DB 463 NEE--MQGLAESIQSMVVERTAKAEVQLQLE-ADMTIKNSIK-----NLQVELATLESTV 515
QY 416 RELEROREERERK-----ELERREAKRELEROROLEMERNNRROELLNQRKEQSTV 469
DB 516 KQLEKQKEATRLRADYDQTOLEESA---CKAQKETKEDTEKMQOIDDQAKAED--- 569
QY 470 LKARKTLFEFLALNDKK--HOLEGKLODIRGLATQROEISTNSRELRLAIEITHQ 527
DB 570 CRA-----NDEKEMELKKEIEMLDNPFKTVRGIVKETSQRQKVAELT--- 614
QY 528 QOLQSQMLGRLIPKQILSDQLKQVQONSILHRDSLTLTKLALAEAKELARQOLR-EOLD 586
DB 615 -----TLERKE-ARDQIQMEKLD 631
QY 587 -EVERETRSKLOEITDVFNNQKLEIREIHSKQOLQKQSLERAKIKQOEKKSILEKQ 645
DB 632 AAIETTKLTQVSDAVEKSEEMVQI-----LRSQQLSTVIDQSLSDDTVYGETAG 686
QY 646 EDQRRVQERDKQWLEHYQOEOPRRPKRPHEDRLKREDSVYRKKAEERAKPEMODKQR 705
DB 687 TSSQ-----NHVQOPDP-----FASARAPAD----- 710
QY 706 LFRPHQEPKATLQAPMSTTEKGLTISAQESVYVYRALYPFESRSHDEITIQPDIV 765
DB 711 -----FPAQVDQ-----FGSSGH----- 723
QY 766 MDESQTEGPGWLGELKGTGTFPAN--YAEKIPENEVPTPAKPVTDLTSAAPKTLAR 823
DB 724 -FDAAFPTDPFAQGG-----FPSDSGFAQSAAPAK-----PAP-----RPAPKSAR 764
QY 824 ETPAPLPTVTSSE-PSTTPNMADFS 847
DB 765 ETPVNDPFAPBSQGSTOPAGFADFA 789

RESULT 11
Q91WH8 PRELIMINARY; PRT; 599 AA.
ID Q91WH8;
AC Q91WH8;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Similar to epidermal growth factor receptor pathway substrate 15,
DE related sequence.
GN Eps15-RS.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC015259; AAH15259.1; -.
DR MGD; MGI:104582; Eps15-rs.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000261; Eps15_repeat.
DR Pfam; PF00036; ehnd; 4.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
KW Receptor.
SQ SEQUENCE 599 AA; 66206 MW; B62762A063F21FA9 CRC64;

Query Match 7.9%; Score 493.5; DB 11; Length 599;
Best Local Similarity 24.8%; Pred. No. 1.2e-15;
Matches 152; Conservative 96; Mismatches 189; Indels 177; Gaps 13;

QY 15 WAITVERAKHDQFLSKPIAGFITGDQARNFFQSGLPQPVLAQIWMALADNMNDGRMD 74

```

```

DB 121 WAVEEKAFCDFEFISLIPVNGILSGDKVPLVMSKSLDLVDLIGVWDISDIDKXGHLID 180
QY 75 QVESIMXKILKLOGYOLPSTLPPVMKQOVALISSAPAFGIGIASMPPLAVAPVPE 134
DB 181 RDEFAVAMHLVYRALKEKPEVPSILPEPL-----IPSKKKTIVA 220
QY 135 GSIPVGMSPPLVSSVPAAVPLPANGAPVIOPLPAFAHPAATWPKSS--SFSRSGPGS 192
DB 221 GAVFVLSPSP-----PDSIRSPSHGVS 246
QY 193 QUNTKLOAKOSFYVASAPPAEAVNPOSSSLKROLFNSHDKTMSGHLTPQARTIMOS 252
DB 247 SLNSTGSLSPKHSV-KQPPVA-WVPVADKRRFDEITFLKTDLDLDGVSQGEVKEIFMHS 304
QY 253 SLPPQALASIMNLSIDODGKLTAEFFILAMHLIDVAMS-GQPLPVPPEVYPPSEFRV 311
DB 305 GLTONLAMIHALADTQGTGLSKGEFALAMVFIQKVSXGIDPQVLSDFWPPSEKGT 364
QY 312 RSGSGMSVSSSVQDRLPEPSSSEDOQPEKKLPVTFEDKRENFERSGVELEKRRQAL 371
DB 365 PIPSSSTLASGEF-----TGVEKLDLDSQ-- 389
QY 372 LEQQRKQERLAOLERARQERKERROEOEAKROLLEKOLEKORERQEREERKEIE 431
DB 390 -----EIAQLOREKYSLEQDIREKEKIKQKTEVQ-----ELQNDLD 427
QY 432 FREAKRELEROROLEWERNRROELNQRNKEQGTVALKARKKTLFELEALNDKKHOL 491
DB 428 RETSLQLEAKQKQDAQR-----IDEMQOQKAKL 457
QY 492 EKGKQDTRCRLATOROELESINKREBLIAITHLOOQLQESQOMGRLLPEKQILSDOL 551
DB 458 RDMLSDVAKQKQDFTQITISLKTOIOSQESDLKQEDLNRKSELNRLQOE---TOL 513
QY 552 KOVQONSILHRSLTLTKALAKELARQOLREQLDEVERETRSLKQIDVFNQKLEIRE 611
DB 514 EQ-----SIQAGRAQJLETLIRSLKCTQODINQARSLQLOE 550
QY 612 IHSKQOOLOKORSLE 625
DB 551 SH-----LEAHRSL 560
RESULT 12
QY 09M11 PRELIMINARY; PRT; 668 AA.
AC 09M11;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE CG16932 protein
GN ERS-15 OR CG16932.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Tracheata; Brachyera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN 111
RC SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman U.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Baeson K.Y., Basso P.V., Berman B.P., Bhandari D., Bolintsov S.,
RA Botkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

```

```

RA Cherry J.M., Cawley S., Dahke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Dubin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glöckner A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.O., Wei M.-H., Idegum C.,
RA Jafarizadeh M., Kallush F., Karpen G.H., Ke Z., Kemson J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laekow P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weisenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
DR EMBL; AE003465; AAF47267.1;
DR Flybase; FBgn0035060; Eps-15.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR00261; Eps15_repeat.
DR Pfam; PF00036; efhand; 4.
DR SMART; SM00054; Efb; 2.
DR SMART; SM00027; EH; 3.
DR PROSITE; PS00018; EF_HAND; 2.
DR Calcium-binding.
SQ
SEQUENCE 668 AA; 73275 MW; 10896D2431A27A2E CRC64;
Query March 7.7%; Score 485; DB 5; Length 668;
Best Local Similarity 25.4%; Pred. No. 3,3e-15;
Matches 162; Conservative 104; Mismatches 224; Indels 148; Gaps 20;
QY 5 PTFPGSLD-VWATVERARHDDQFLSKPLIAGFTTGDQARNFFPQGLDQPLAQIWA 63
DB 116 PVASGVANGDWSIGVIDRLKYEQLFESIHPSNGMLPNKVKYGVLMDSKLPMSILGTIWD 175
QY 64 LADNMNDGMDQVESIMXKILKLOGYOLPSTLPPVMKQOVALISSAPAFGIGIASM 123
DB 176 LADQDQDGLDHEFVAMHLVYOTLQKRTIPSVLPBLKRP-----GAGP 222
QY 124 PLTVAVAPVPMGSIPIVGMSPPLVSS-----VPPAAVPLANGAPPVI 166
DB 223 PPKPAMPPEPPACA-----AMPBAPSGFGDGFVANFPKDIAPPAIPLPVAVPWMT 276
QY 167 QPLPAFAHFAATWPKSSFSRSGPSQUNTKLOAKOSFYVASAP--PAEAVNPOSSSLK 224
DB 277 RIPPV-----GAVSSQPLIQDPLPIGAPMANAVWVPADIKR 317
QY 225 YROLFNSHDKTMSGHLTPQARTIMOSLSLPPQALASIMNLSIDODGKLTAEFFILAMH 284
DB 318 FEFIFQSDLDGDLGVSGLGVKDIIFKSGIIPORSLADIALCTNOSKLTVEQFALAMW 377
QY 285 LIDVAMSGQPLPVLPEVYPPSEFRVRSVSGMSVSSSVQDRLPEPSSSEDOQPEKK 344
DB 378 FVERKQRGVDPHVLNANVPPSMKATVAGVDL-----QPEV 415
QY 345 LPVTFEDKRENFERSGVELEKRRQALLEQQRKQERLAOLERARQERKER-EROEOEAK 403
DB 416 KP-TYSPLEMIISKEIELARER-----RVLEETIAKQADVAIKGESLD 462
QY 404 RQLELEKQERLELERQEREERKEIERREA-----KRELEROOLEWERNRROE 455
DB 463 TLTATIKQJLENGRGEAKRLDQLQOVSHNTAVLANVSLDSRTNEQVTKIRDCQHQEV 522

```

```

QY 456 LNMORKEOGTVLVKARRKTELELEALNDKKHOLEGKLODIRCLATQROESTNKS 515
DB 523 TIN-----EOGE--LNAKRS-----ELÖKLDKEASLOKEYDSNNRBLSTLNLHÖAT--- 569
QY 516 RELRIAEITHLOOQLOESQOMLGRILPEKQILSDOLKOVQNSLHDSLLTLKRALE--- 572
DB 570 -OQJSSVRSMVTÖLETÖROM-----TDALLICRAAMENON 605
QY 573 AKELARQOLREOLD-EVERETRSK---LOEIDVF--NN 604
DB 606 AELVSEYÖLKIEPDPDEARKTILKEVÖLPKDDPEENN 643

RESULT 13
Q9HGL2 PRELIMINARY; PRT; 1116 AA.
ID 09HGL2;
AC 09HGL2;
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Hypothetical 120.5 kDa protein.
GN SPBC800.10C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA McCombie W.R., Lyne M.;
RT "Sequence analysis of a region of the fission yeast genome.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA McCombie W., Lodhi M., Kaplan N., Johnson A.;
RT "Sequence of a region of the fission yeast genome.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL391034; CAC01525.1; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR00261; EFS1_repeat.
DR Pfam; PF00036; ehand; 4.
DR SMART; SM00054; EFh; 3.
DR SMART; SM00027; EH; 3.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
KM Hypothetical protein.
SQ SEQUENCE 1116 AA; 120540 MW; 79F1941C4CEBAF9 CRC64;

Query Match 6.9%; Score 434.5; DB 3; Length 1116;
Best Local Similarity 21.0%; Pred. No. 1.4e-12;
Matches 215; Conservative 152; Mismatches 320; Indels 339; Gaps 39;

```

```

QY 301 -----PEYIP-----PSFRVRSQGSVSISSSVYD----- 326
DB 355 VPTSPISVVPANISSPNPNPTLAPNPTGSPSVTSGTEDLSDATPPSPPLAPOHTSSNA 414
QY 327 -----ORLPPESSDEDO-----PEKLPVTFEDKKREN 356
DB 415 TKHSAPVTKSAPPVVSPILOHNTPGFPTSPPAKPNSPTSTFPQSSFGOTIAKNTMDK 474
QY 357 FERGSELEKEROALIEQORKEORLEAOLEPAPEREREREROEOEAROLEKOR 416
DB 475 PSAVTSPSQLAAPIPVASAE---QKLAAVPRKESOLQYKKSNDLÖK---SSR 527
QY 417 ELERQEEERREKIERREPAKREOROLEWERNROELNORKEOGTVLVKARKT 476
DB 528 DVAAVLSDVKAKVSEIRAYDEELAKAKÖI-----S 558
QY 477 LEFELEALNDKKHOLEGKLODIRCLATQROESTNKSRELRIAEITHLOOQLOESQOM 536
DB 559 LDIE-----TNKAÖQVVRREYSILBATNALÖKÖKÖKGEV 595
QY 537 LGRLPEKQILSDOLKOVQNSLHDSLLTLKRALLEAKELARQOLREOLD-EVERETRSKL 596
DB 556 LEQVVAE-----SEAAKNVSS--NASTIÖLKSEVADKEÖTLAÖHLÖLDEMTÖ---RL 645
QY 597 ÖEIDVFNNÖKLEBEIHSKÖÖLÖKÖRSLEAARLÖKÖERKSLELEKÖK---EDAQRV 652
DB 646 VSLDESKAVSQ-RKLDLEYKINNSKTÖLATATBEYHSHSQLEAKÖELSKLEDGLKSV 704
QY 653 ÖERDÖKÖLEHVÖEÖPPEPRKPHEDRLKEDSVRKKEAEERAKPEMODKÖSLFHH-- 710
DB 705 N-----LTBEAPKP---EVDSTPRPSP--TSNGITTDKPLPDITSSVPTÖHNS 749
QY 711 -----ÖEPAKLA--TQAPWSTTEKCP---LTISA-----QESVKVYVYALY 747
DB 750 FDMAMNTLRSPSLNSNNSAASIVSRNPFNLKISGASSPSVSNFWSSESPASAVFPSPIS 809
QY 748 PFESRSHDEITÖPÖ-----DIVMD---ESÖTGEPGMLGELGKT 786
DB 810 KTTLSLVNNSVNSVNSLOSEPVOGLNMEPÖHQDSVVDVTSASÖQRSP--VLSDLÖKLT 867
QY 787 GWFPAVNAEKIPENEVPTPAKPVTDLTSAPAP-KLALRE-----TPAPL 829
DB 868 G-SANTAEAPV-EN---TSAPPIEN--TSAPPEIANKQÖATPEISAPFATERISTAPV 921
QY 830 --PVTSSE-----PSTPPNMADEFSSTW-----PSSNEKEPDTDMÖTMAÖP 870
DB 922 KPVPSPSRDRSAÖDGVVQÖATPHIÖDEFPPIÖPNEIDDESSDEEPMSN---LSP 976
QY 871 SLTVPS 876
DB 977 QISÖS 982

RESULT 14
Q18137 PRELIMINARY; PRT; 944 AA.
ID Q18137;
AC Q18137;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE Hypothetical 113.6 kDa protein.
GN C25A11.4B.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL_N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
"Genome sequence of the nematode C. elegans: a platform for

```



QY 387 RAQERKERERQO-----EAKQULELEKOLEKOLERQERERKEIERREA 436  
 Db 333 RLIRQOREKREEREDRLBSIRLAEBAELARRALEKER-IDREKAEERKTMERLERE 391  
 QY 437 KRELEOROLEWERNRRO---ELNQRNKEOGTV---VLKARKTLEFELALNDKCH 489  
 Db 392 RARLERER-LEERRROKEREKETERIERERREHRIEIERIKRIRIEREREREKKA 450  
 QY 490 QLEGLQDIRCLATROEIESTNKS-RELRIAEITHLOOQLOESQOMUGRLIPEKQILS 548  
 Db 451 EED-----RLRRLERLERIERERRELEAREQOLEIQRRREADR-----ERQRL 496  
 QY 549 DQUKQY-QQNSLHRDULITIKRLEKELARQOLEQDEVERETSKLOEIDVFNQUL- 606  
 Db 497 DEAREMRRREREERREAEADVHRQAEERERLKKQREBEARLERIRLEQOKIDMERID 556  
 QY 607 KELREITHSKOOLQOKORSL-EAARLKQEOERKSL---ELEKOKEDARVRVOPRDKOWLEH 662  
 Db 557 AERREERERKEERREFEELI EAARRKKEARDRDLDEMERERVAREERERKEERQERRI 616  
 QY 663 VQOEOPRPKPKPHE-----DRLKREDSYRKKEAEERAKPEMO 700  
 Db 617 AAEKERKROEEREEELIARLNELQRAAARQAQONNAELDRQORQDELDRK-AQELSEREMR 675  
 QY 701 DKQSLFHPHOEPAKLATQAPWSTTEKGPLTISAQESVKVYTRALYPFESRSHDEITQ 760  
 Db 676 EKERRDRERANBEAQLADLIERERHNQULIRENERREAVERRANNRLE--DRSRDCL--- 730  
 QY 761 PGDIVWVDESQTEGPEWLGELGKGTGWFPANYAEKIPENEVPTPAKPYTDLTSAPAKL 820  
 Db 731 --DHIVERSEKEQ-----FELEKRLLAEKEANNRKK--NHLISSE---- 768  
 QY 821 ALRETPAPLPVTSSEPSTP-----NNWADFSSTWPSSS 854  
 Db 769 TLAKLTQPMVYTTREPEVTTKVERQYIERIDRYWVEDVPYAPSQS 814

Search completed: March 14, 2003, 12:09:44  
 Job time : 51 secs



